~~H05 0077~~
Thesis/
Reports
Murphy,
A. H.

5120

Statistics and Climatic Impacts Laboratory (SCIL)

Department of Atmospheric Sciences

SCIL Report 84-2
(USDA-FS-RMFRES 28-K3-302)

THE STATE OF THE ART OF PROBABILISTIC WEATHER FORECASTING

Allan H. Murphy

Barbara G. Brown

August 1984

Oregon State University

Corvallis, Oregon 97331

Statistics and Climatic Impacts Laboratory (SCIL) Reports

This series of reports will contain scientific papers, progress and project reports, and other special reports in the areas of statistical and applied meteorology and climatology prepared by members of the SCIL staff and/or by individuals associated with SCIL projects. These reports are intended to make the results of SCIL activities more widely available and easier to reference. Copies of the reports will be provided on a first-come, first-serve basis until supplies are exhausted.

Allan H. Murphy
Director, SCIL

Note

This report (SCIL 84-2) represents the second in a series of reports prepared under a project sponsored by the Rocky Mountain Forest and Range Experiment Station, Forest Service, U.S. Department of Agriculture.

THE STATE OF THE ART OF PROBABILISTIC WEATHER FORECASTING

Report #2

to

Rocky Mountain Forest and Range Experiment Station
Forest Service, USDA
Fort Collins, Colorado 80528

by

Allan H. Murphy and Barbara G. Brown[1]

Department of Atmospheric Sciences
Oregon State University
Corvallis, Oregon 97331

August 1984

---

[1]Current address: Environmental and Societal Impacts Group, National Center for Atmospheric Research, Boulder, Colorado 80307.

Abstract

This report very briefly describes the rationale for quantifying the uncertainty in weather forecasts and the status of probability forecasting in meteorology. A short review of the existing - and quite modest - literature on probabilistic fire-weather forecasting is presented and the literature on probability forecasting in meteorology is introduced. An extensive paper summarizing the state of the art of probabilistic weather forecasting is included as an appendix.

Contents

Section Number and Title		Page
1.	Introduction	1
2.	Probabilistic Weather Forecasting: Rationale and Status	2
3.	Probabilistic Fire-Weather Forecasting: A Brief Review	4
4.	Probability Forecasting in Meteorology: An Introduction to the Literature	6
	References	9
	Appendix	11

1. Introduction

This report is one of a series of reports prepared under a project sponsored by the Rocky Mountain Forest and Range Experiment Station of the Forest Service, U.S. Department of Agriculture. The overall objective of the project was "to develop procedures to quantify the uncertainty in spot fire-weather forecasts." However, the specific goals of the project included the preparation of a review of the state of the art of probabilistic weather forecasting. Fortunately, an extensive review paper on this topic had recently been completed by the principal investigator, and this paper appears as an appendix to the present report.

The rationale for quantifying the uncertainty in weather forecasts and the status of probability forecasting in meteorology are very briefly summarized in Section 2. Section 3 contains a short review of the existing - and quite modest - literature on probabilistic fire-weather forecasting. An introduction to the literature on probability forecasting in meteorology is provided in Section 4. As indicated above, the state of the art review of probabilistic weather forecasting is included in an appendix.

2. Probabilistic Weather Forecasting: Rationale and Status

Weather forecasts are inherently uncertain. This uncertainty exists whether the forecasts are based on numerical (i.e., physical/dynamical) and/or statistical models or on the subjective judgments of human forecasters. In the context of weather forecasting, the sources of uncertainty include an incomplete description of the state of the atmosphere at any point in time, assumptions and approximations made in formulating mathematical and conceptual models of the atmosphere, and necessary simplifications in the procedures used to derive forecasts from these models.

The basic rationale for quantifying the uncertainty in weather forecasts resides in the fact that this information is needed by users of the forecasts to make optimal decisions in uncertain situations. In fact, it is relatively easy to show that the economic value of probability forecasts generally exceeds the economic value of categorical (i.e., nonprobabilistic) forecasts for all users of such forecasts (e.g., see Murphy, 1977; Krzysztofowicz, 1983). Thus, the formulation and utilization of probability forecasts provides a means of enhancing the usefulness of weather forecasts - by improving the effectiveness of users' decision-making procedures - without any advance in the state of the art of weather forecasting.

Considerable experience has been gained in quantifying the uncertainty in weather forecasts in recent years, especially in the United States. Objective numerical/statistical procedures now routinely provide probability forecasts for a variety of weather elements, and these forecasts are made available to National Weather Service (NWS) forecasters as guidance and to other individuals who subscribe to the NWS communication system (see Glahn, 1985; Murphy, 1985). Moreover, subjective precipitation probability forecasts have been prepared on an operational, nationwide basis for almost

20 years, and subjective probability forecasts of many other weather events have been formulated on an experimental basis in recent years (see Hughes, 1980; Murphy, 1985). In most cases, these objective and subjective probability forecasts have been found to be both reasonably reliable and quite skillful.

3. Probabilistic Fire-Weather Forecasting: A Brief Review

Fire behavior is very sensitive to meteorological conditions, especially temperature, relative humidity, precipitation, and wind. Moreover, it is widely recognized and generally agreed that the effective use of forecasts of these weather elements offers a means of improving decisions related to fire management (e.g., Furman, 1982: Radloff and Yancik, 1983). Rational and optimal use of weather forecasts in fire management - as well as in many other decision-making problems - requires quantitative information concerning the uncertainty in the forecasts. Thus, it is somewhat surprising to discover that very few studies of probability forecasts have been conducted in the context of fire-weather forecasting and that the uncertainty in fire-weather forecasts is generally not expressed in quantitative terms on an operational basis.

The authors have been able to locate only four papers or reports concerned with probability forecasting in the context of forest or fire management (Ayer, 1968; Furman, 1982; Schroeder, 1954; Turner, 1959). We briefly review these publications in chronological order in this section. The paper by Schroeder (1954) summarizes the results of an experiment in which probabilistic terms (e.g., probable, possible, chance of) were incorporated into fire-weather forecasts of precipitation occurrence formulated in the Chicago District in 1952. Evaluation of the experimental results indicated that the forecasters possessed some "skill" in assessing the probability of occurrence of precipitation, especially for lead times less than 30 hours. The paper by Turner (1959) reports the results of an experiment in which numerical probabilities were assigned to the occurrence of measurable precipitation at eleven locations in British Columbia by forecasters at the Vancouver Public Weather Office. As in the case of the

study for the Chicago District, the forecasters in Vancouver exhibited significant skill in assigning probabilities to such precipitation events.

In a technical memorandum, Ayer (1968) discussed several issues related to the interpretation, verification, and use of probability forecasts in the context of fire-weather forecasting. The author focused specifically on forecasts of precipitation events and lightning storms. He also made several recommendations regarding probability forecasting in the Portland Fire-Weather District, including proposals involving the formulation of experimental probability forecasts of the relevant events.

The use of probabilistic weather information in decision-making problems involving prescribed fire is discussed by Furman (1982). He investigated and compared the value of different types of information, including climatological probabilities, persistence forecasts, subjective categorical forecasts, and forecasts involving a combination of persistence and subjective forecasts. The latter three types of information were actually expressed as conditional probabilities (these probabilities were derived by calibrating the forecasts on the basis of their performance). The combined forecasts were found to be of greatest value. In conclusion, the author indicated that "an analytical framework will help the decision maker derive maximum utility from the uncertain weather forecasts available."

Although we have not uncovered any other directly relevant publications, such publications may exist. Moreover, probability forecasts are undoubtedly formulated from time to time in conjunction with some fire-weather forecasting programs. In this regard, objective and subjective probabilistic forecasts prepared in the context of public and other general weather forecasting programs (e.g., probability of precipitation forecasts) may be routinely incorporated into fire-weather forecasts for some locations.

4. Probability Forecasting in Meteorology: An Introduction to the Literature

The literature on probabilistic weather forecasting, which was quite modest until 1965, has grown rapidly in the last two decades. For the purposes of this introduction to the literature, it is convenient to divide the subject into four specific topics: (a) rationale and basis for probability forecasting; (b) formulation of probability forecasts by objective and subjective methods; (c) verification of probability forecasts (methods and results); and (d) use and value of probability forecasts. Several papers and reports have treated at least some aspects of all of these topics; however, most publications have focused on a specific topic. A reasonably comprehensive bibliography on all aspects of probabilistic weather forecasting up to the late 1960s was prepared by Murphy and Allen (1970). In this section we will focus on more recent contributions to the subject.

To our knowledge, no in-depth treatment of the rationale and basis for probabilistic weather forecasting has appeared in the meteorological literature. However, authors of several papers and/or reports have briefly discussed various facets of this topic. A recent treatment can be found in Sections 1 and 2 of the chapter by Murphy (1985) (see Appendix).

Uncertainty in weather forecasts can be quantified objectively or subjectively (or by a combination of objective and subjective methods). Objective probability forecasting is based on the use of statistical models or procedures, and recent reviews of this topic include the chapters by Glahn (1985) and Murphy (1985) and the review paper by Murphy and Winkler (1984). Meteorological journals (especially Monthly Weather Review and the Journal of Applied Meteorology - the latter is now the Journal of Climate and Applied Meteorology) and selected conference preprint volumes (especially the

American Meteorological Society's series in Weather Forecasting and in Probability and Statistics) contain numerous applications of such methodology to objective probability forecasting.

Subjective probability forecasting has recently been reviewed by Murphy (1985) and Murphy and Winkler (1984). Operational subjective probability forecasting programs now exist in the U.S. and in several other countries; these programs generally involve probabilistic forecasts of precipitation occurrence. Moreover, numerous experiments involving subjective probability forecasts for a variety of weather variables have been conducted since 1970. Subjective probability forecasts are also formulated on a regular basis by participants in forecasting games associated with synoptic laboratory programs at several universities (e.g., see Bosart, 1975; Sanders, 1979).

A variety of methods have been developed to verify probability forecasts. Some of these procedures and verification measures are described in a recent review paper by Murphy and Daan (1985). Other papers of general interest on this topic include Murphy (1976), Murphy and Epstein (1967), and Winkler and Murphy (1968).

The use and value of probability forecasts is a topic with several different facets, including the interpretation and understanding of such forecasts. An important but frequently overlooked paper on this topic is that by Roberts (1968). The optimal use and economic value of probability forecasts has been studied from several different perspectives by a number of individuals, including Krzystofowicz (1983), Murphy (1977), Stuart (1982), Thompson (1962), and Winkler and Murphy (1979). Considerable controversy has arisen concerning public interpretation and understanding of probability forecasts, and this issue has been addressed recently by Murphy *et al.* (1980).

Finally, practical problems frequently arise in conjunction with the design and/or implementation of probability forecasting programs (and, for that matter, any new forecasting program). Some of these problems are discussed in reports by the Air Weather Service (1978) and Hughes (1980). Moreover, Murphy and Winkler (1979) outline some steps that should be taken prior to the implementation of an operational probability forecasting program.

References

Air Weather Service, 1978: Probability forecasting: a guide for forecasters and staff weather officers. Scott Air Force Base, Illinois, U.S. Air Force, Headquarters Air Weather Service, Pamphlet No. 105-51, seven chapters and attachments (paged separately).

Ayer, H.S., 1968: Probability forecasting in the Portland Fire-Weather District. Salt Lake City, Utah, ESSA, Weather Bureau, Technical Memorandum WBTM WR-32, 18 pp.

Bosart, L.F., 1975: SUNYA experimental results in forecasting daily temperature and precipitation. Monthly Weather Review, 103, 1013-1020.

Furman, R.W., 1982: The effectiveness of weather forecasts in decision making: an example. Journal of Applied Meteorology, 21, 532-536.

Glahn, H.R., 1985: Statistical weather forecasting. Probability, Statistics, and Decision Making in the Atmospheric Sciences (A.H. Murphy and R.W. Katz, Editors). Boulder, Colorado, Westview Press, pp. 289-335.

Hughes, L.A., 1980: Probability forecasting - reasons, procedures, problems. Silver Spring, Maryland, NOAA, National Weather Service, Technical Memorandum NWS FCST 24, 84 pp.

Krzysztofowicz, R., 1983: Why should a forecaster and a decision maker use Bayes' theorem. Water Resources Research, 19, 327-336.

Murphy, A.H., 1976: Evaluation of probabilistic forecasts: some procedures and practices. Weather Forecasting and Weather Forecasts: Models, Systems, and Users - Volume 2 (A.H. Murphy and D.L. Williamson, Editors). Boulder, Colorado, National Center for Atmospheric Research, pp. 807-830.

Murphy, A.H., 1977: The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situation. Monthly Weather Review, 105, 803-816.

Murphy, A.H., 1985: Probabilistic weather forecasting. Probability, Statistics, and Decision Making in the Atmospheric Sciences (A.H. Murphy and R.W. Katz, Editors). Boulder, Colorado, Westview Press, pp. 337-377.

Murphy, A.H., and R.A. Allen, 1970: Probabilistic prediction in meteorology: a bibliography. Silver Spring, Maryland, ESSA, Weather Bureau, ESSA Technical Memorandum WBTM TDL 35, 60 pp.

Murphy A.H., and H. Daan, 1985: Forecast evaluation. Probability, Statistics, and Decision Making in the Atmospheric Sciences (A.H. Murphy and R.W. Katz, Editors). Boulder, Colorado, Westview Press, pp. 379-437.

Murphy, A.H., and E.S. Epstein, 1967: Verification of probabilistic predictions: a brief review. Journal of Applied Meteorology, 6, 748-755.

Murphy, A.H., and R.L. Winkler, 1979: Probabilistic temperature forecasts: the case for an operational program. Bulletin of the American Meteorological Society, 60, 12-19.

Murphy, A.H., and R.L. Winkler, 1984: Probability forecasting in meteorology. Journal of the American Statistical Association, 79, 489-500.

Murphy, A.H., S. Lichtenstein, B. Fischhoff, and R.L. Winkler, 1980: Misinterpretations of precipitation probability forecasts. Bulletin of the American Meteorological Society, 61, 695-701.

Radloff, D.L., and R.F. Yancik, 1983: Decision analysis of prescribed burning. Preprints, Seventh Conference on Fire and Forest Meteorology (Fort Collins). Boston, Massachusetts, American Meteorological Society, pp. 85-89.

Roberts, H.V., 1968: On the meaning of the probability of rain. Proceedings, First National Conference on Statistical Meteorology (Hartford). Boston, Massachusetts, American Meteorological Society, pp. 133-141.

Sanders, F., 1979: Trends in skill of daily forecasts of temperature and precipitation, 1966-78. Bulletin of the American Meteorological Society, 60, 763-769.

Schroeder, M.J., 1954: Verification of "probability" fire-weather forecasts. Monthly Weather Review, 82, 257-260.

Stuart, A., 1982: On the economic value of probability of precipitation forecasts in Canada. Journal of Applied Meteorology, 21, 495-498.

Thompson, J.C., 1962: Economic gains from scientific advances and operational improvements in meteorological prediction. Journal of Applied Meteorology, 1, 13-17.

Turner, J.A., 1959: The evaluation of rainfall probability forecasts for forestry. Toronto, Canada, Department of Transport, Meteorological Branch, CIR-3148, TEC-290, 9 pp.

Winkler, R.L., and A.H. Murphy, 1968: "Good" probability assessors. Journal of Applied Meteorology, 7, 751-758.

Winkler, R.L., and A.H. Murphy, 1979: The value of weather forecasts in the cost-loss ratio situation: an ex ante approach. Preprints, Sixth Conference on Probability and Statistics in Atmospheric Sciences (Banff). Boston, Massachusetts, American Meteorological Society, pp. 134-138.

Appendix

This appendix contains a copy of a paper entitled "Probabilistic Weather Forecasting" by A.H. Murphy that appeared as Chapter 9 in Probability, Statistics, and Decision Making in the Atmospheric Sciences (A.H. Murphy and R.W. Katz, Editors), published by Westview Press (Boulder, Colorado) in 1985.

CHAPTER 9

PROBABILISTIC WEATHER FORECASTING

Allan H. Murphy

1. INTRODUCTION

Weather forecasts are inherently uncertain. This uncertainty is present whether the forecasts are based on "objective" numerical (i.e., physical/dynamical) and/or statistical models, on the subjective judgments of human forecasters, or on a combination of objective and subjective procedures. Uncertainty in the weather forecasting process arises from several sources. First, observational systems and networks in meteorology provide, at best, an incomplete description of the state of the atmosphere at any point in time. Second, the models used to describe the behavior and evolution of the atmosphere - and its attendant weather - necessarily involve a variety of assumptions and approximations concerning the relevant physical and dynamical processes. Third, the procedures used to derive forecasts from these models generally entail additional simplifications or approximations. These factors all make significant contributions to the uncertainty that currently resides in weather forecasts.

The importance of quantifying the uncertainty in weather forecasts and of communicating this information to users of forecasts was first recognized more than seventy-five years ago (Cooke, 1906a,b). In the intervening period, a variety of studies and experiments in probabilistic weather forecasting have been undertaken (see Sections 3 and 4; see also Murphy and Winkler, 1984). This work has included the development and application of statistical and numerical-statistical methods of probability forecasting as well as the formulation of probability forecasts by weather forecasters on an operational and experimental basis. A significant event in the historical development of probabilistic weather forecasting occurred in 1965, when the National Weather Service (NWS) in the United States initiated a nationwide operational program of subjective precipitation probability forecasting. Moreover, for more than a decade now, NWS forecasters have received objective precipitation probability forecasts as guidance in the formulation of their subjective forecasts. In recent years, subjective and objective probability forecasts of other meteorological events have been prepared on an experimental and/or

operational basis in the U.S. and probabilistic weather forecasting has been initiated in several other countries.

This chapter is concerned primarily with methods of formulating probability forecasts and with the results of operational and experimental probability forecasting programs in meteorology. First, the definition and interpretation of probability forecasts and the motivation for probability forecasting are discussed briefly in Section 2. Section 3 treats objective probability forecasts, including methods of preparing such forecasts as well as some results of current objective probability forecasting programs. Subjective probability forecasts are considered in Section 4, with subsections containing an introduction to methods of probability assessment and aggregation and the results of operational and experimental subjective probability forecasting programs. Section 5 examines some issues related to the communication of uncertainty in weather forecasts provided to the general public and specific users.

2. PROBABILITY FORECASTS: DEFINITION, INTERPRETATION, AND MOTIVATION

A probability forecast can be considered to be a discrete or continuous probability distribution defined over the values of the predictand of interest. It will be assumed that such distributions satisfy the usual probability axioms (e.g., Lindley, 1965, pp. 6-14). In the case of a discrete predictand with N possible values or classes, a probability forecast consists of a set of N probabilities determining a discrete distribution. Since the probabilities must sum to one, only N-1 probabilities need to be specified. A probability forecast for a continuous predictand consists, in general, of a probability density or distribution function. Such a function might be specified in parametric form (e.g., by indicating that the forecast follows a normal or beta distribution) or graphically (by drawing the density or distribution function by hand). Probabilistic weather forecasts for continuous predictands generally are described in terms of certain summary measures of the underlying distributions (e.g., means, variances, credible intervals). In this chapter, we will be concerned primarily with probability forecasts for discrete predictands.

As indicated in Section 1, probability forecasts can be formulated on the basis of objective models or procedures or on the basis of subjective judgments of forecasters. In the case of the former, the probabilities generally are given a relative frequency interpretation and are considered to be estimates of the "true" (but unknown) probabilities of the events of interest. On the other hand, when probability forecasts are formulated on a subjective basis the probabilities can be interpreted as expressions of a forecaster's subjective degrees of belief concerning the likelihood of occurrence of the relevant events. In this framework it is not necessary (or even

appropriate) to postulate the existence of true probabilities and, as a result, such forecasts frequently are referred to as probability assessments (as opposed to probability estimates). It is important to recognize that a sound theoretical basis exists for both the objective (i.e., relative frequency) and subjective (i.e., personal) interpretations of probability (e.g., Lindley, 1965; Winkler, 1972). For the purposes of this chapter, the differences between these two interpretations of probabilities will not be of major importance.

Two principal reasons exist for formulating and expressing weather forecasts in probabilistic terms. First, as noted in Section 1, uncertainties are inherent in the forecasting process, whether this process is based on objective procedures or subjective judgments. The existence of these uncertainties implies that categorical or deterministic forecasts will seldom if ever be warranted. In other words, such forecasts will seldom accurately reflect the true "state of knowledge" of the forecasting system (regardless of whether the system consists of an objective model or a human forecaster). Only a forecast expressed in probabilistic terms can provide an appropriate description of this state of knowledge. Second, in the absence of perfect forecasts, users require estimates of the likelihood of occurrence of the relevant events in order to make optimal decisions in uncertain situations. It is relatively easy to demonstrate that the value of probabilistic forecasts generally is equal to or greater than the value of categorical forecasts for all users of such forecasts (e.g., Thompson, 1962; Murphy, 1977; Krzysztofowicz, 1983). Thus, a strong motivation exists for formulating and expressing weather forecasts in probabilistic terms.

3. OBJECTIVE PROBABILITY FORECASTS

As noted in Chapter 8, an objective weather forecasting system is a system that produces one and only one forecast from a given set of data. Thus, an objective forecast does not depend directly on the judgment and/or experience of a particular meteorologist, although subjective judgment plays an important role in the development of objective forecasting systems. In Section 3.1 we briefly describe several methods of objective forecasting that can be used to obtain probabilistic forecasts, and in Section 3.2 we present some results of recent operational objective probability forecasting programs in the U.S.

3.1 Methods of Objective Probability Forecasting

Objective probabilistic weather forecasts have been prepared on an experimental and/or operational basis using a variety of statistical, numerical-statistical, and other procedures. Many of these procedures are described in some detail in Chapter 8. Thus, the

discussion of methods of objective forecasting in this section will be relatively brief and will focus on the use of such methods to obtain probability forecasts. References are provided for those readers who seek more detailed descriptions of the procedures themselves or examples of their application.

Contingency tables represent perhaps the simplest procedure for obtaining estimates of the probability of occurrence of future weather events. In this procedure the ranges of values of the predictand and each predictor are divided into two or more mutually exclusive and collectively exhaustive categories or classes, and the data available are used to obtain a discrete joint frequency distribution over the cells in the table. Conditional probability estimates (i.e., forecasts) for the predictand classes can then be obtained from this joint distribution and the relevant marginal relative frequencies. Two-way tables involving the predictand and a single predictor represent the simplest form of this procedure and have been used most frequently, but n-way ($n > 2$) or multidimensional tables also have been developed in some studies. A problem that frequently arises in connection with the use of contingency tables is that the number of cases in some cells may be quite small, thereby leading to unstable probability estimates. This problem necessarily becomes more acute as the number of predictors (i.e., the dimensions of the table) is increased. To reduce the effects of small sample sizes, it may be necessary to limit the number of predictors, to combine classes of the predictand and/or predictors, or to smooth the probability estimates. The contingency table procedure is quite flexible and does not require the analyst to make any assumptions concerning the form of the joint distribution of the variables involved. The work of Wahl et al. (1952) contains examples of the use of contingency tables to obtain objective probability forecasts. In a related vein, Gringorten (1966, 1971), Martin (1972), and others have developed simple conditional probability models that can be used in a manner similar to contingency tables to produce probabilistic forecasts on the basis of climatological data.

Scatter diagrams, or graphical regression, can also be used to obtain probability forecasts of events of interest. This procedure has been described in some detail in Chapter 8 and that description will not be repeated here. It is sufficient to indicate that the procedure can accommodate several predictors and provides the meteorologist or forecaster with a considerable degree of flexibility in analyzing relationships between variables (although care must be taken not to "over-analyze" the data). As noted in Chapter 8, it is a particularly suitable procedure for small data samples and generally has been based on a subjective analysis process, although efforts have been made to automate the procedure (e.g., see Freeman, 1961). Scatter diagrams were studied extensively by Brier (1946), who used them to obtain probability forecasts of precipitation amounts. Other examples of the application of this procedure to derive probability forecasts of various events include papers by Dickey (1949), Thompson (1950), and Schmidt (1951).

A variety of statistical procedures (or models) can be used to produce objective probability forecasts, and several of these procedures were described in some detail in Chapter 8. The model employed most frequently in objective statistical or numerical-statistical weather forecasting is multiple linear regression. Multiple regression is used to obtain probability forecasts by treating the predictand as a set of two or more Boolean (or dummy) variables. This variant of standard regression was developed and popularized in the meteorological community by Miller (1964), and it is usually referred to as Regression Estimation of Event Probabilities (REEP). As noted in Chapter 8, the REEP procedure has the important property that it minimizes the Brier score, a mean square error measure of the accuracy of probabilistic forecasts (see Chapter 10). The probabilities generated by the REEP model for a set of mutually exclusive and collectively exhaustive events must sum to one, but they are not constrained to the closed interval [0,1] (this latter problem is not serious and can be readily rectified). In meteorological applications, the REEP procedure generally is used in conjuction with a forward stepwise technique for predictor selection, and the predictors themselves frequently are transformed into Boolean variables. Extensive use of the REEP model is made in the current NWS operational objective weather forecasting program (see Section 4.2; see also Chapter 8).

Multiple discriminant analysis (MDA) is another statistical model that has been used to produce probabilistic forecasts of a variety of weather events. As indicated in Chapter 8, this procedure is particularly appropriate for nominal (or non-numerical) predictands, and it is also able to incorporate certain nonlinear relationships between the predictors and the predictand. A nonparametric approach to discriminant analysis, based on the work of Fix and Hodges (1951), was used by Miller (1962) to obtain probability forecasts for ceiling height and visibility events. This work also involved the development and application of a forward stepwise technique for predictor selection. Extensive use was made of the MDA procedure in a variety of objective probabilistic weather forecasting studies at the Travelers Research Center in Hartford, CT in the 1960's (e.g., see Enger et al., 1964). Subsequent comparisons of the MDA and REEP models revealed that MDA performed somewhat better than REEP for some predictands, but the former imposed a greater computational burden than the latter.

Other objective procedures that have been used to generate probabilistic weather forecasts on an experimental and/or operational basis include the logit model (Brelsford and Jones, 1967; Jones, 1968), adaptive logic (Glahn, 1964), and analogues (Kruizinga and Murphy, 1983). The logit model is currently used operationally to produce probability forecasts of precipitation type (see Section 3.2).

The objective procedures described in this section (and in Chapter 8) generally can be used in conjunction with any of three approaches to statistical or numerical-statistical weather forecasting. These approaches are the classical method, the perfect prog procedure, and the model output statistics (MOS) technique, and they are described and compared in detail in Chapter 8. Probabilistic forecasts can be generated using any of these approaches in combination with a suitable statistical procedure. Comparative evaluation of objective forecasts for a variety of weather elements indicates that forecasts based on the MOS approach usually outperform forecasts based on the other two approaches (see Chapter 8).

Another objective method of obtaining estimates of the uncertainty inherent in weather forecasts is stochastic-dynamic prediction (Epstein, 1969). This method involves the use of dynamical equations to describe the evolution of the variances and covariances (and possibly higher moments of the probability distributions) of the variables of interest - as well as the use of an equation for the evolution of the mean - and uncertainty is introduced by means of the initial variances and covariances. To date, only relatively simple dynamical models have been employed in conjunction with the stochastic-dynamic approach. The use of this approach together with the more sophisticated models currently employed in operational numerical weather prediction would impose a severe computational burden on the forecasting system, and this fact alone precludes the practical application of the stochastic-dynamic method at this time.

An extensive operational program of objective statistical weather forecasting based primarily on the MOS approach now exists in the NWS. As noted in Chapter 8, approximately 90,000 regression and logit equations are used daily to produce nearly 600,000 forecasts. A substantial fraction of these forecasts are formulated in probabilistic terms and many of these probabilities are provided to the NWS forecasters as guidance and to subscribers to the NWS communications system.

3.2 Some Recent Operational Results

In this section we briefly describe the current NWS operational objective probability forecasting program in the U.S. and present some recent results of this program. At the end of the section, objective probabilistic weather forecasting in other countries will be mentioned. In evaluating the results presented here, emphasis will be placed on the reliability and skill of the forecasts. For a detailed discussion of methods of evaluating probability forecasts, see Chapter 10.

As indicated in Section 3.1, the NWS objective probability forecasting program is based largely on the MOS approach, in which statistical techniques such as multiple regression analysis are

employed to develop prediction equations relating numerical model output (and other data) and the weather variables of interest. This approach is used to produce probabilistic weather forecasts on an operational basis for a variety of weather variables, and these forecasts are provided to NWS forecasters as guidance. A list of the predictands for these probabilistic forecasts is presented in Table 1. Information regarding predictand element, number of categories or thresholds, category/threshold definitions, forecasts periods (where appropriate), and lead times is included in this table. The contents of the table demonstrate the wide range of predictands for which objective probabilistic guidance forecasts are now prepared by the NWS. In this section we present and briefly describe some results of the MOS probability forecasting program for five predictands; namely, precipitation occurrence, precipitation type, cloud amount, ceiling height, and visibility.

Objective probabilistic forecasts of precipitation occurrence have been prepared on an operational basis using a regression model in conjunction with the MOS approach since 1972 (Lowry and Glahn, 1976), and these forecasts generally are referred to as objective probability of precipitation (PoP) forecasts. Some recent results of the NWS objective PoP program are presented in Figure 1 and Table 2. Figure 1 contains reliability diagrams for these forecasts for three periods with lead times of 12-24 hours, 24-36 hours, and 36-48 hours, respectively. The data represent the 1980 warm season (April-September) and the 1980-1981 cool season (October-March) and consist of the composite results for 87 stations in the conterminous U.S. In a reliability diagram, the relative frequency of occurrency of the event of interest is plotted against the forecast probability for specific probability values (in this case, the thirteen permissible probability values are 0.00, 0.02, 0.05, 0.10, 0.20, ..., 1.00). The diagonal 45° line represents perfect reliability, in the sense that relative frequency is exactly equal to forecast probability (see Chapter 10 for further discussion of reliability). The reliability curves in Figure 1 indicate that the correspondence between forecast probabilities and observed relative frequencies is quite good for both seasons and all three periods. The insets in the reliability diagrams depict the relative frequency of use of the probability values. The frequency of use distributions for the three periods are similar, with the extreme probabilities (i.e., 0.00 and 1.00) being used more frequently in the cool season than in the warm season.

Average Brier scores and skill scores for these samples of objective PoP forecasts are presented in Table 2. The average Brier score is the mean square error of the forecasts and the skill score is the percent improvement in the average Brier score for the forecasts over the average Brier score for forecasts based solely on climatological probabilities (see Chapter 10 for formal definitions of these evaluation measures). The scores in Table 2 indicate that, as expected, the accuracy (Brier score) and skill (skill score) of the forecasts decrease as the lead time increases. Moreover, the overall

Table 1. Predictands for which objective probabilistic forecasts are prepared on an operational basis by the U.S. National Weather Service.

Predictand element[a]	Number of categories (C)/ thresholds (T)	Category/ threshold definitions	Forecast periods (hours)[b]	Lead times (hours)[c]
Precipitation occurrence	1(T)	≥ 0.01 inches	6,12	6-12,12-18,...,54-60 12-24,24-36,...,48-60
Precipitation amount	4(T)	≥ 0.25; ≥ 0.50; ≥ 1.00; ≥ 2.00 inches	6,12	6-12,12-18,...,36-42 12-24,24-36,36-48
Type of liquid precipitation	3(C)	drizzle; rain; showers		18,30,42
Precipitation type	3(C)	snow; freezing rain; rain		6,12,...,60
Snow amount	3(T)	≥ 2.0; ≥ 4.0; ≥ 6.0 inches	12	12-24
Cloud amount	4(C)	clear; scattered; broken; overcast		6,12,...,60
Ceiling height	6(C)	< 200; 200-400; 500-900; 1000-2900; 3000-7500; > 7500 feet		6,12,...,48
Visibility	6(C)	< 1/2; 1/2-7/8; 1-2 3/4; 3-4; 5-6; > 6 miles		6,12,...,48
Obstructions to vision	4(C)	none; haze/smoke; blowing phenomena; fog/ground fog		6,12,...,48
Thunderstorms	1(T)	VIP[d] level ≥ 3	4,12,24	2-6,12-24,24-36, 36-48,12-36
Severe local storms (watches)	1(T)	tornadoes, hail, and/or strong winds	4	2-6
Severe local storms (outlooks)	1(T)	tornadoes, hail, and/or strong winds	12,24	12-24,24-36, 36-48,12-36
Convective wind gusts	1(T)	thunderstorm wind gusts ≥ 25 knots	3	8-15,20-28
Dew (ground condensation) intensity	3(C)	light; moderate; heavy		36,60,84

[a]Forecasts of the type of liquid precipitation, precipitation type, and severe local storms (outlooks) are expressed in the form of conditional probabilities. Forecasts of all other predictand elements are expressed in the form of unconditional probabilities.

[b]Predictand elements with unspecified forecast periods (e.g., cloud amount) are valid at specific times (see entry under Lead times).

[c]Lead times are defined with respect to the latest numerical model output used as input in preparing the forecasts (this time is taken as the zero point).

[d]VIP = Radar Video Integrator and Processor.

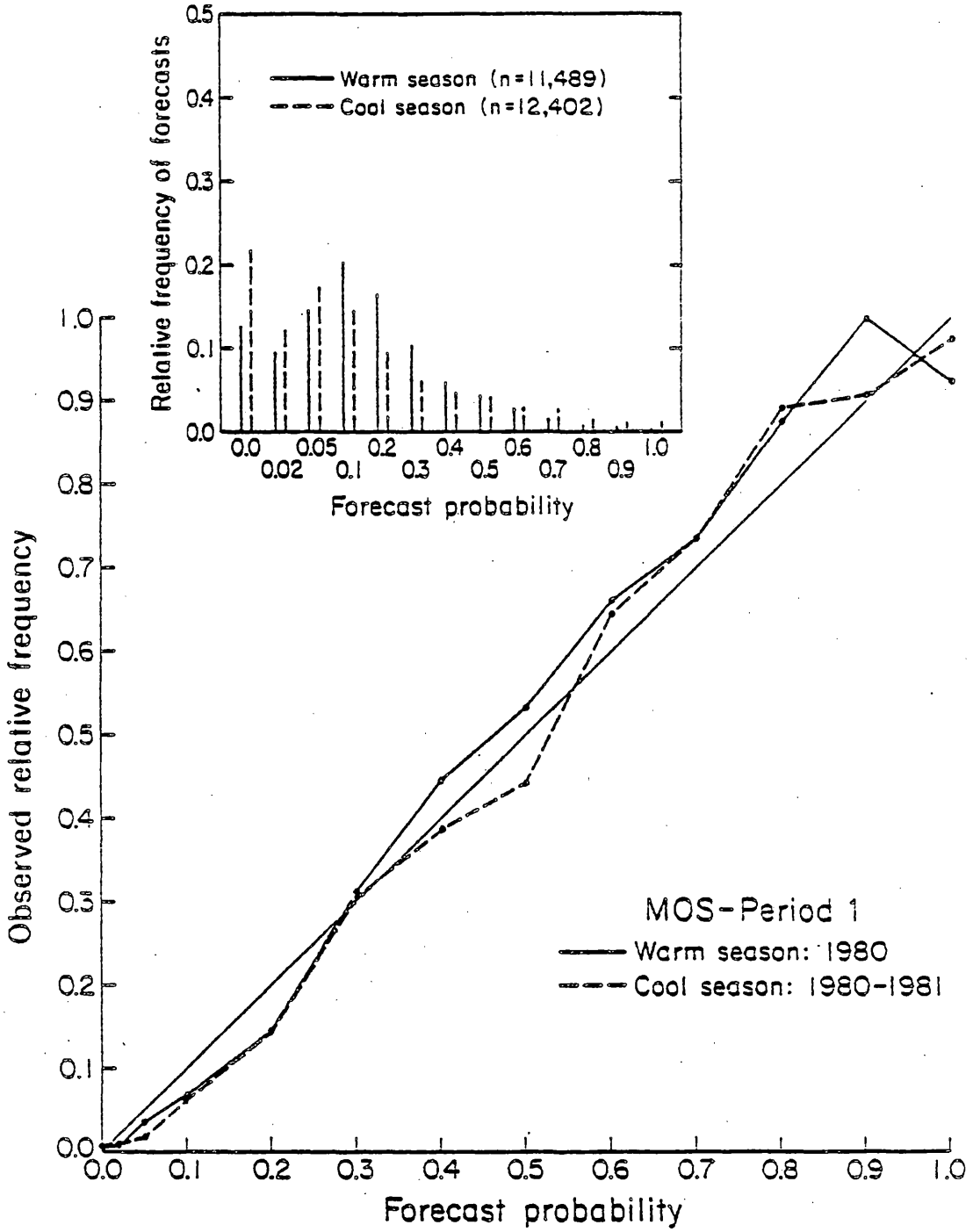

Figure 1a. Reliability diagram for first-period (12-24 hours) objective PoP forecasts in 1980 warm season and 1980-1981 cool season for approximately 85 stations in conterminous U.S. Inset indicates relative frequency of use of probability values in respective seasons.

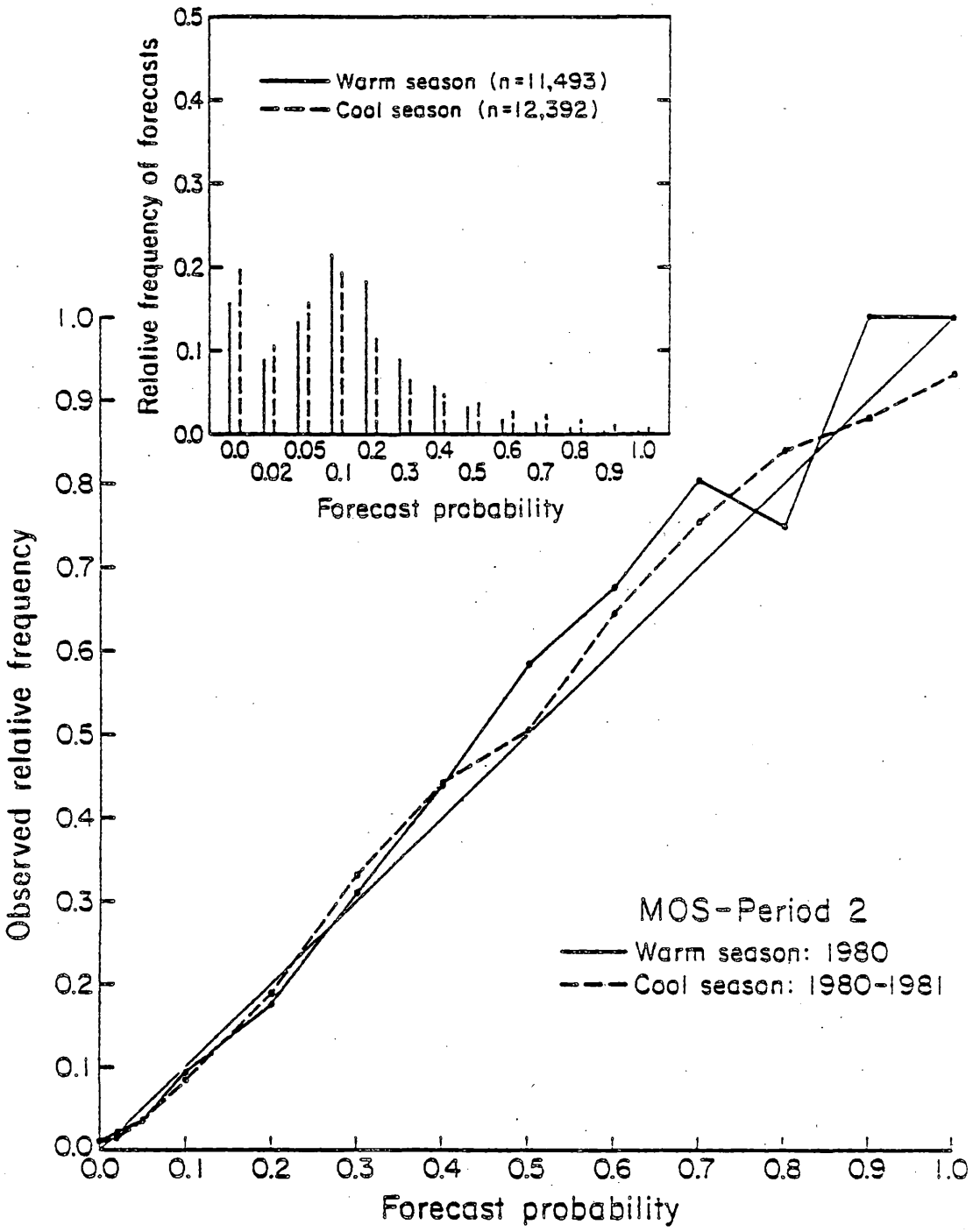

Figure 1b. Reliability diagram for second-period (24-36 hours) objective PoP forecasts in 1980 warm season and 1980-1981 cool season for approximately 85 stations in conterminous U.S. Inset indicates relative frequency of use of probability values in respective seasons.

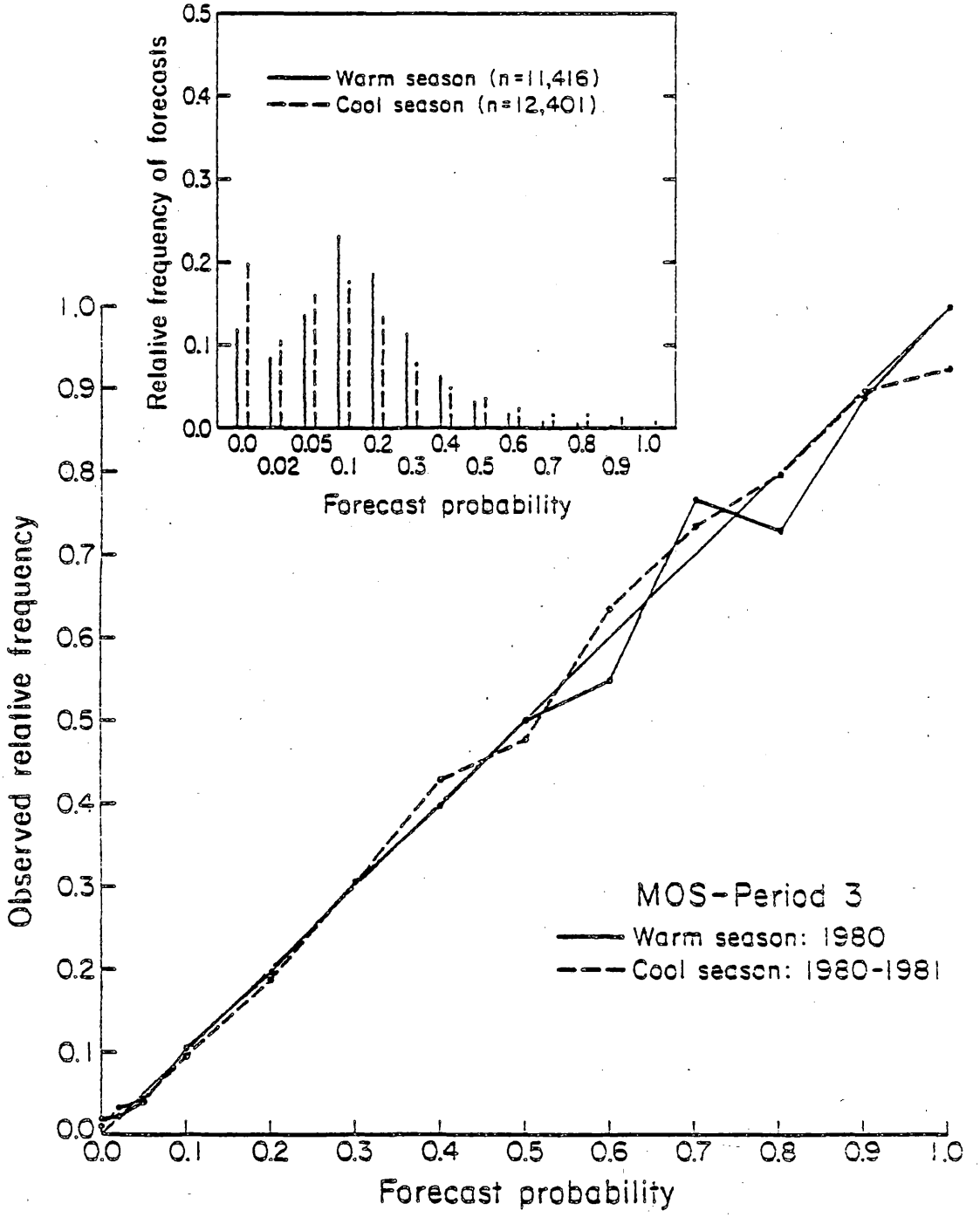

Figure 1c. Reliability diagram for third-period (36-48 hours) objective PoP forecasts in 1980 warm season and 1980-1981 cool season for approximately 85 stations in conterminous U.S. Inset indicates relative frequency of use of probability values in respective seasons.

Table 2. Average Brier scores and skill scores based on Brier score of MOS probability of precipitation forecasts for three 12-hour periods (initial time 0000 GMT) in 1980 warm season and 1980-1981 cool season for approximately 85 stations in the conterminous U.S. (Maglaras et al., 1981; Schwartz et al., 1981).

	Warm season: 1980			Cool season: 1980-1981		
Period	Number of forecasts	Average Brier score[a]	Skill score (%)	Number of forecasts	Average Brier score[a]	Skill score (%)
1	11,489	0.0993	26.5	12,402	0.0772	46.8
2	11,493	0.1023	21.8	12,392	0.0912	34.2
3	11,416	0.1144	14.7	12,401	0.0959	33.2

[a]Average Brier score reported here is one-half of original Brier score.

accuracy and skill of the objective PoP forecasts are considerably better in the cool season than in the warm season (presumably due primarily to the fact that the cool season, unlike the warm season, is dominated by large-scale weather systems that are quite well predicted by numerical models).

Conditional objective probabilistic forecasts of precipitation type in three categories (snow or sleet, freezing rain, rain) are prepared operationally by the NWS using a logit model in conjunction with the MOS approach (Bocchieri, 1979a). The forecasts are conditional in the sense that the probabilities indicate the likelihoods of occurrence of the three types of precipitation given that measurable precipitation occurs (i.e., the forecasts are not "in force" if precipitation does not occur). Average Brier scores and skill scores for a dependent sample of MOS precipitation type forecasts are presented in Table 3 for four lead times. These data represent the "winter" season (September-April) and consist of the composite results for 15 stations in the conterminous U.S. (see Table 3 for additional details). The scores indicate that these forecasts are quite skillful at all lead times. It should be noted that skill scores for an independent (e.g., operational) sample of objective forecasts generally would be somewhat lower than the scores reported in Table 3. For a more detailed description of this program and some additional results, including reliability diagrams, see Bocchieri (1979a).

A regression model is also used in conjunction with the MOS approach to produce objective probability forecasts of cloud amount, ceiling height, and visibility (Globokar, 1974; Carter and Glahn, 1976). These forecasts are expressed in the form of discrete probability distributions for four (cloud amount) and six (ceiling height, visibility) categories of the relevant variables (see Table 1). Some results of this operational program from the 1980-1981 cool season (October-March) for nine stations in the southeastern U.S. are presented in Table 4. In addition to average Brier scores and the associated skill scores, average ranked probability scores and the corresponding skill scores are also given for these forecasts. The latter is a quadratic scoring rule that is particularly appropriate for the evaluation of probability forecasts of multiple-category ordinal variables (see Chapter 10 for a description of the ranked probability score and its properties). These scores indicate that the 6-hour and 12-hour forecasts for the three elements all possess positive skill and that, as expected, skill decreases as lead time increases. Since the ranked probability score (unlike the Brier score) is sensitive to distance, the differences between the values of the two skill scores suggest that the MOS probability forecasts of these elements are quite successful in assigning high probabilities to the observed category or to categories close to the observed category (this attribute of the forecasts is not necessarily reflected in the Brier score).

Table 3. Average Brier scores and skill scores based on Brier score of conditional probability forecasts of precipitation type for four lead times (initial time 0000 GMT) in "winter" season (September-April) for 15 stations in the conterminous U.S.[a] Results based on developmental sample for nine years (1972-1981) for 12-hour and 24-hour forecasts and for five years (1976-1981) for 36-hour and 48-hour forecasts.

Lead time (hours)	Number of forecasts	Average Brier score	Skill score[b] (%)
12	3,533	0.1193	71.5
24	3,028	0.1329	67.7
36	2,159	0.1668	60.6
48	1,795	0.1796	56.8

[a]Stations include Albany, NY; New York, NY; Norfolk, VA; Charleston, WV; Atlanta, GA; Flint, MI; St. Louis, MO; Little Rock, AK; Minneapolis, MN; Grand Island, NE; Williston, ND; Amarillo, TX; Spokane, WA; Salt Lake City, UT; Winslow, AR.

[b]Skill score is based on climatological probabilities for developmental sample for winter season at each station.

Table 4. Average Brier and ranked probability scores and corresponding skill scores for six-hour forecasts of selected elements for 1980-1981 cool season at 233 stations in the conterminous U.S.

Element	Number of forecasts	Brier score		Ranked probability score	
		Average score	Skill score[a] (%)	Average score	Skill score[a] (%)
Cloud amount	36,387	0.4087	41.5	0.1006	85.7
Ceiling height	36,391	0.2972	42.2	0.0445	91.0
Visibility	36,434	0.1828	32.7	0.0342	89.2
Obstructions to vision	36,436	0.1082	50.4	0.0453	84.4

[a]Skill scores based on sample climatological probabilites for all stations combined for 1980-1981 cool season.

In addition to probability forecasts of precipitation occurrence, precipitation type, cloud amount, ceiling height, and visibility, the MOS approach currently is used to produce objective probabilistic forecasts of a variety of other elements on an operational basis (see Table 1). To conserve space, results of these other programs are not presented here. Instead, we provide a list of some references that the reader can study to find specific results of the element(s) of interest: precipitation amount (Bermowitz and Zurndorfer, 1979); snow amount (Bocchieri, 1979b); severe local storms (watches) (Charba, 1979); severe local storms (outlooks) (Reap and Foster, 1979); and dew (Jensenius and Carter, 1979). It should be noted that the probability forecasts produced by the MOS system for many of these variables frequently are converted into categorical forecasts either for evaluation purposes or when they are disseminated via the NWS communications system. Thus, it is sometimes difficult to find results in the published literature concerning the quality of these objective forecasts in their original probabilistic form. General references containing results of the MOS program, as well as comparisons of objective and subjective forecasts, include Zurndorfer et al. (1979), Charba and Klein (1980), Glahn (1980), and Chapter 8 in this book.

In recent years, objective probabilistic weather forecasting has also been undertaken on an operational and/or experimental basis in several other countries. For example, operational precipitation probability forecasts have been made in France since 1979 using a perfect prog approach (Javelle et al., 1980); extensive tests of an analogue/perfect prog approach to formulating probability forecasts of precipitation occurrence and precipitation amount have been conducted in Canada (Wilson and Yacowar, 1980); and a procedure based on decision trees has been used on an operational basis for thunderstorm probability forecasting in the German Democratic Republic since 1979 (Balzer, 1980). In addition, many other "small" studies of objective probability forecasting have been conducted on an experimental basis (e.g., Mason, 1977).

4. SUBJECTIVE PROBABILITY FORECASTS

Subjective probabilistic weather forecasts have been formulated on an experimental and/or operational basis for many years (see Murphy and Winkler, 1984). By subjective probability forecasts, we mean probability forecasts whose formulation depends ultimately on the real-time judgments of one or more human forecasters. Of course, objective guidance information, including objective probabilistic weather forecasts, may play an important role in the process of preparing these subjective forecasts. In any case, according to this definition, the probability of precipitation (PoP) forecasts currently disseminated to the general public in the U.S. are subjective forecasts.

Historically, subjective probability forecasting in meteorology has been conducted largely in an intuitive manner. That is, formal methods of assessing and/or aggregating probabilities have seldom been used. Nevertheless, we believe that such procedures are quite useful and that they should be employed more widely in the context of weather forecasting. Thus, a brief discussion of these assessment and aggregation procedures appears in Section 4.1. Some results of recent operational and experimental subjective probability forecasting programs are presented in Section 4.2.

4.1 Assessment and Aggregation of Subjective Probabilities

In formulating weather forecasts, forecasters generally assimilate information in a largely intuitive manner from a variety of sources. Here, we assume that this assessment and assimilation process leads to the formulation of internal judgments that represent the forecasters' true beliefs concerning the likelihoods (i.e., probabilities) of occurrence of the relevant events. Thus, weather forecasts stated or recorded by forecasters represent "externalizations" of these judgments, and the former may not always correspond to the latter. In this regard, since judgments are "expressed" in probabilistic terms, only probability forecasts can correspond exactly to the judgments. Close correspondence between judgments and forecasts obviously is quite desirable and this consideration plays an important role in both probability assessment and evaluation (e.g., see Winkler and Murphy, 1968; see also Chapter 10 in this book). Here, we discuss some methods of probability assessment that can help to ensure good correspondence between judgments and forecasts and then briefly consider the problem of the aggregation of probabilistic assessments.

As noted earlier, weather forecasters generally have made their probability assessments (i.e., forecasts) in an intuitive manner. That is, they have seldom used any of the formal procedures that are available to assist individuals in making subjective probability forecasts. Descriptions of some of these procedures can be found in the literature of statistics, operations research, and management science (e.g., Winkler, 1967; Staël von Holstein, 1970; Hampton et al., 1973; Spetzler and Staël von Holstein, 1975). Probability assessment procedures can be classified in different ways; here, we follow the scheme outlined in Spetzler and Staël von Holstein (1975). Most of these procedures are based on questions (posed by an interviewer or the assessor himself/herself) for which the answers can be represented as points on a cumulative distribution function. One dimension of this classification scheme relates to whether the assessor (or forecaster) is asked to assign probabilities (with the values of the variable fixed) or to assign values (with the probabilities fixed). To illustrate this difference, the fixed-width and variable-width credible interval temperature forecasts formulated in several recent experiments (e.g., Murphy and Winkler, 1974; Winkler and Murphy, 1979) represent examples of assigning probabilities (with values fixed) and

of assigning values (with probabilities fixed), respectively. The second dimension of the classification scheme relates to whether the forecaster's response mode is direct or indirect. A direct response mode requires numbers as answers (to the questions), whereas an indirect response mode requires the assessor to choose between one or more bets or alternatives. In the case of the latter the bets are adjusted until the assessor is indifferent in choosing between them, and this indifference can be translated into a probability or value assignment. For example, if a forecaster is indifferent between a payoff of $3.00 for certain and a payoff of $10.00 if measurable precipitation occurs today, then his/her probability of the event "precipitation" on this occasion is 0.3. In the indirect response mode, it is possible to use either internal or external (reference) events. The probability wheel (Spetzler and Staël von Holstein, 1975) represents an example of a device that is useful in specifying external reference events. For further details concerning this two-dimensional classification scheme and for descriptions of probability assessment procedures corresponding to each combination of assignment method and response mode, see Spetzler and Staël von Holstein (1975). We briefly describe a fixed-value, indirect response mode procedure that has been used successfully in meteorological and other contexts in the next paragraph.

A formal procedure that is useful for assessing probabilities of (essentially) continuous variables is the interval technique or "method of successive subdivisions" (Raiffa, 1968). In this technique, the forecaster is required to make a series of equal-odds indifference judgments, each of which divides an interval of values of the variable into two equally likely subintervals. The first judgment leads to the identification of the median (i.e., the value that the forecaster feels is equally likely to be exceeded or not exceeded), and subsequent judgments specify the 0.25 and 0.75 fractiles (lower and upper quartiles), the 0.125 and 0.875 fractiles, etc. Then, 50%, 75%, etc., central credible intervals can be obtained from the values (of the variable) corresponding to these fractiles. Consistency checks can be incorporated into this procedure to ensure that the assessments actually reflect a forecaster's true judgments (e.g., values of the variable falling inside the 75% credible interval should be three times as likely as values falling outside the interval). Dialogues describing the application of this procedure can be found in Raiffa (1968, pp. 161-168), Peterson et al. (1972), and Air Weather Service (1978). The method of successive subdivisions has been applied successfully to variable-width (fixed-probability) credible interval temperature forecasting in several experiments (Peterson et al., 1972; Murphy and Winkler, 1974; Winkler and Murphy, 1979), and some results of these programs are presented in Section 4.2.

In the process of formulating a probability forecast, forecasters assimilate or aggregate information from a variety of sources. That is, they are continually revising their probabilistic judgments concerning the events of interest on the basis of new or additional

information. Weather forecasters generally perform this aggregation process in an informal, intuitive manner, and the question arises as to whether this process is being conducted in an efficient and effective manner. In this regard, formal models exist for the optimal aggregation of information in probabilistic prediction (e.g., Winkler and Murphy, 1973). In brief, these models involve the revision of probabilities in the light of new information within the framework of Bayes' theorem, and they can treat both conditionally independent and dependent information sources (however, the latter may involve complex probability assessments). Although relatively little experience has been gained in the use of such models in real-world (as opposed to laboratory) settings, they offer some promise as techniques for improving the process of information aggregation in subjective (probabilistic) weather forecasting. In a related vein, simple ad hoc procedures, such as hierarchical models (e.g., see Murphy and Winkler, 1971) or decision trees (e.g., Belville and Johnson, 1982), that can be used to decompose the assessment and assimilation processes may be of some value in this context.

If subjective probability forecasts of the same event are made by more than one forecaster, then it is possible to combine these forecasts to form a "consensus" forecast. Considerable evidence now exists in the context of probabilistic weather forecasting to indicate that consensus forecasts formulated simply by averaging the individual probabilities generally will perform as well as most if not all of the forecasters (e.g., Sanders, 1963, 1973; Bosart, 1975). Winkler et al. (1977) compared various procedures for combining individual subjective PoP forecasts, including schemes that ignore the forecasters' past performance and schemes that consider their past performance. These consensus procedures all performed about equally well, and they all performed better than almost all of the forecasters. Thus, when several forecasters are available, it may be possible to improve probability forecasts by combining the forecasts of individual forecasters (of course, one or more of these "individuals" could represent objective forecasting procedures).

It should be noted that subjective probability assessments frequently suffer from certain biases or systematic errors (Tversky and Kahneman, 1974). These biases may be due to the inherent limitations of humans to assimilate information and quantify judgments and/or to their use of heuristics (i.e., simple rules or procedures) to simplify these difficult and complex tasks. Studies of subjective probability assessments made in other contexts generally reveal substantial overconfidence on the part of the forecasters (e.g., Fischhoff and MacGregor, 1982; Lichtenstein et al., 1982; Wallsten and Budescu, 1983). Although weather forecasters frequently represent an exception to this general result, overforecasting has been noted in several experimental probability forecasting programs (e.g., Daan and Murphy, 1982; Murphy et al., 1982; Murphy and Winkler, 1982). This overforecasting usually has been attributed to the forecasters' lack of experience in probability forecasting, to the failure to provide

the forecasters with feedback regarding their performance, and/or to a value-induced bias associated with the impacts of the events and forecasts of interest on the users. A recent experimental study of subjective probabilistic weather forecasting in a "laboratory" setting (Allen, 1982) investigated several possible explanations for overforecasting and other biases. In any case, it is certainly desirable to give forecasters some training in probability assessment prior to the start of any experimental or operational probability forecasting program and to provide them with feedback concerning their performance on a regular basis. Recently, some efforts have been made to provide weather forecasters with background information and training materials in the context of probability forecasting (e.g., Air Weather Service, 1978; Hughes, 1980).

4.2 Some Recent Operational and Experimental Results

Several operational and experimental subjective probability forecasting programs are briefly described in this section. First, recent results of the NWS operational precipitation probability forecasting program are presented and references to similar operational programs in other countries are given. Then, some results of probabilistic temperature forecasting experiments and of a recent probability forecasting experiment in The Netherlands are summarized. Finally, several other experimental subjective probability forecasting programs are mentioned.

As noted in Section 1, subjective PoP forecasts have been prepared operationally on a nationwide basis in the U.S. since 1965. Some recent results of this program are presented in Figure 2 and Table 5. These results relate to the same lead times, seasons, and stations as the results for the objective PoP forecasts described in Section 3.2. Figure 2 contains reliability diagrams for the subjective PoP forecasts for the three lead times. The diagrams indicate that these forecasts are quite reliable in both seasons (warm/cool) for all three periods. Comparison of the reliability diagrams for the SUB and MOS forecasts (cf. Figures 1 and 2) reveals only small differences in reliability. In this regard, it should be noted that the reliability of the SUB forecasts is not due solely to the reliability of the MOS forecasts, since the subjective PoP forecasts formulated by NWS forecasters prior to the availability of the objective MOS forecasts were also very reliable (e.g., Roberts et al., 1969). The frequency of use distributions (insets) are similar for the three periods, although a slight tendency exists for the extreme probabilities to be used less frequently as lead time increases. Comparison of these distributions with the corresponding distributions for the objective PoP forecasts indicates that extreme probabilities are assigned to this event (i.e., the occurrence of measurable precipitation) more often by the forecasters than by the MOS system.

Average Brier scores and skill scores for the subjective PoP

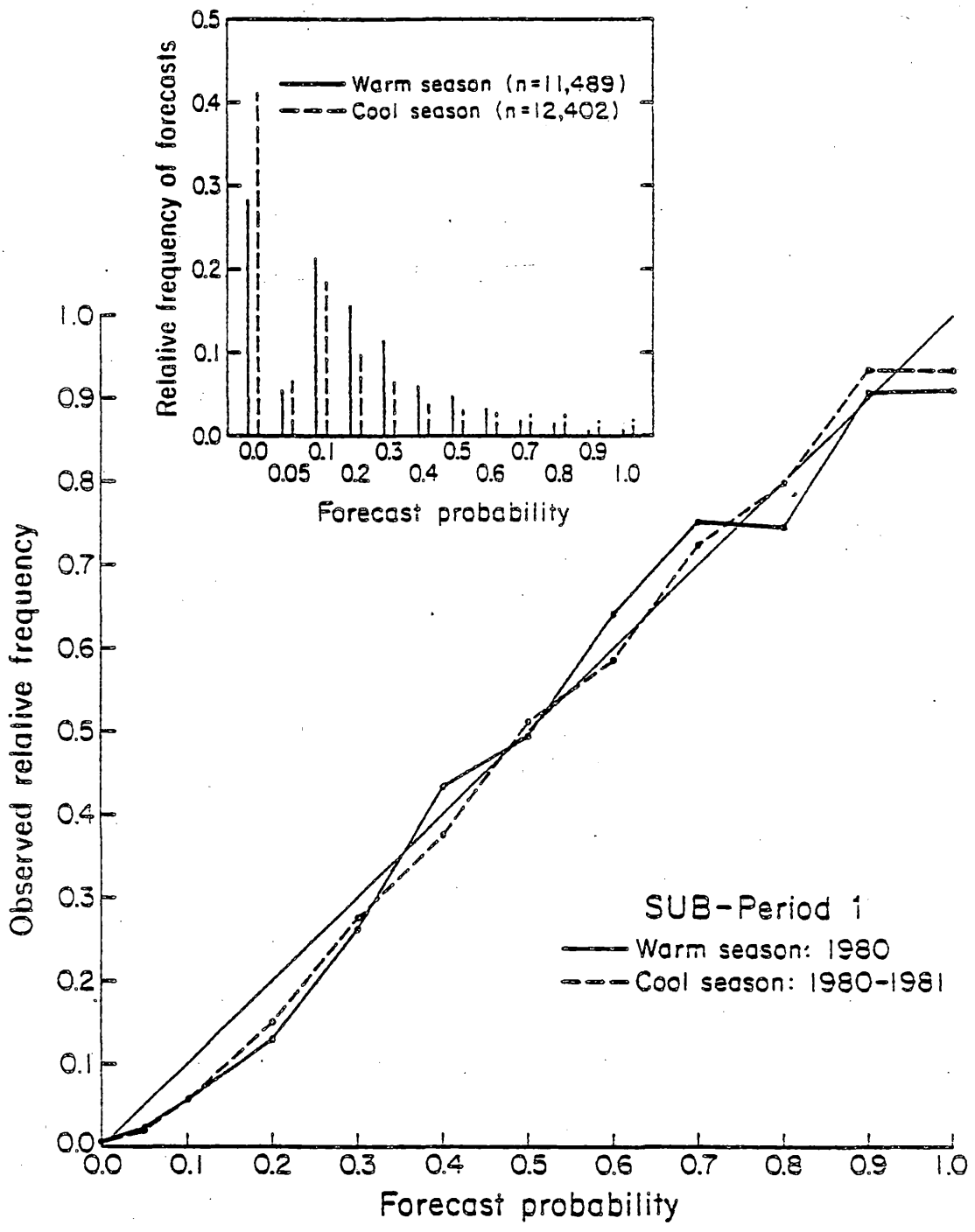

Figure 2a. Reliability diagram for first-period (12-24 hours) subjective PoP forecasts in 1980 warm season and 1980-1981 cool season for approximately 85 stations in conterminous U.S. Inset indicates relative frequency of use of probability values in respective seasons.

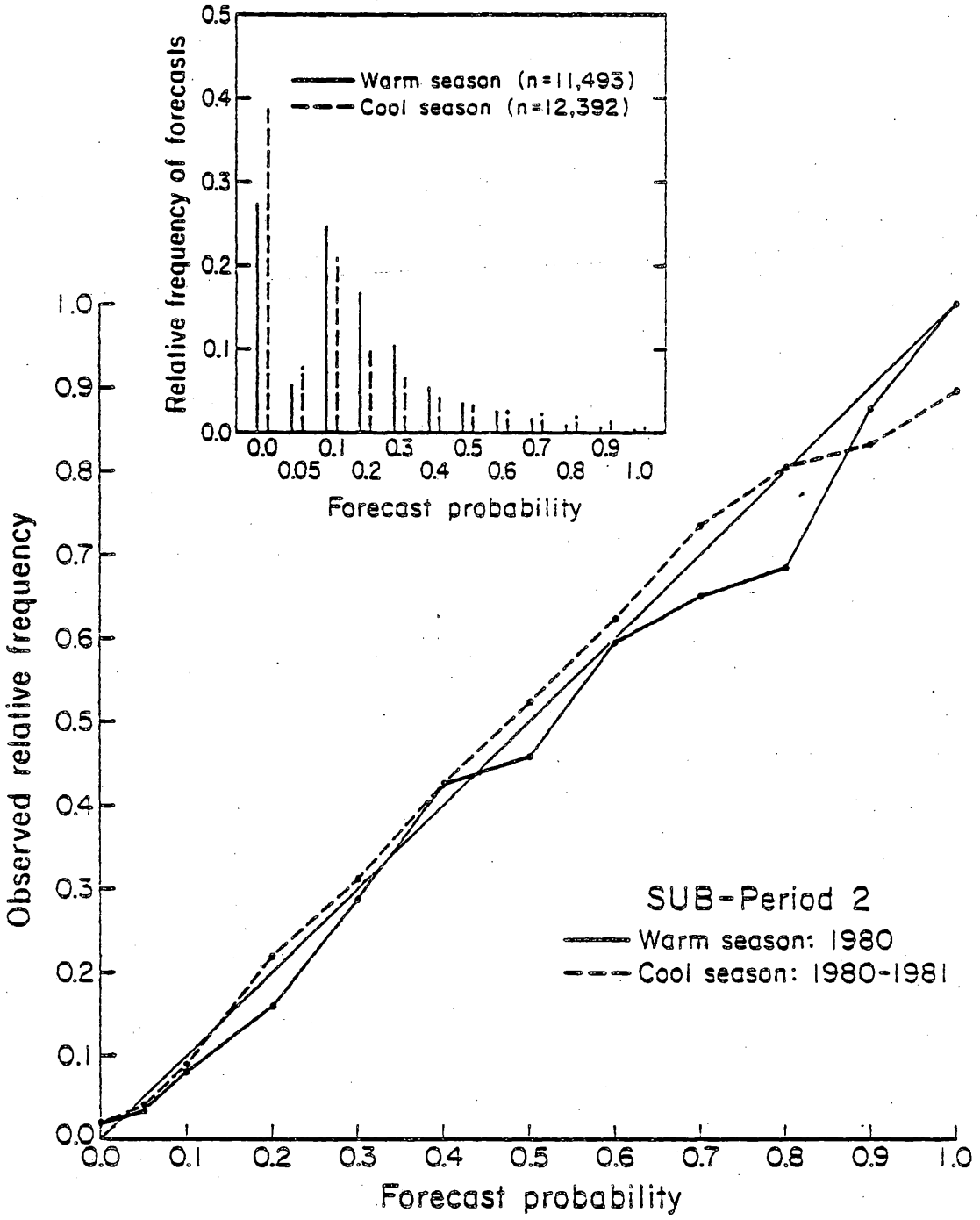
Relative frequency of forecasts
Warm season (n=11,493)
Cool season (n=12,392)
0.0 0.1 0.2 0.3 0.4 0.5
0.0 0.05 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Forecast probability
Observed relative frequency
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Forecast probability
SUB-Period 2
Warm season: 1980
Cool season: 1980-1981

Figure 2b. Reliability diagram for second-period (24-36 hours) subjective PoP forecasts in 1980 warm season and 1980-1981 cool season for approximately 85 stations in conterminous U.S. Inset indicates relative frequency of use of probability values in respective seasons.

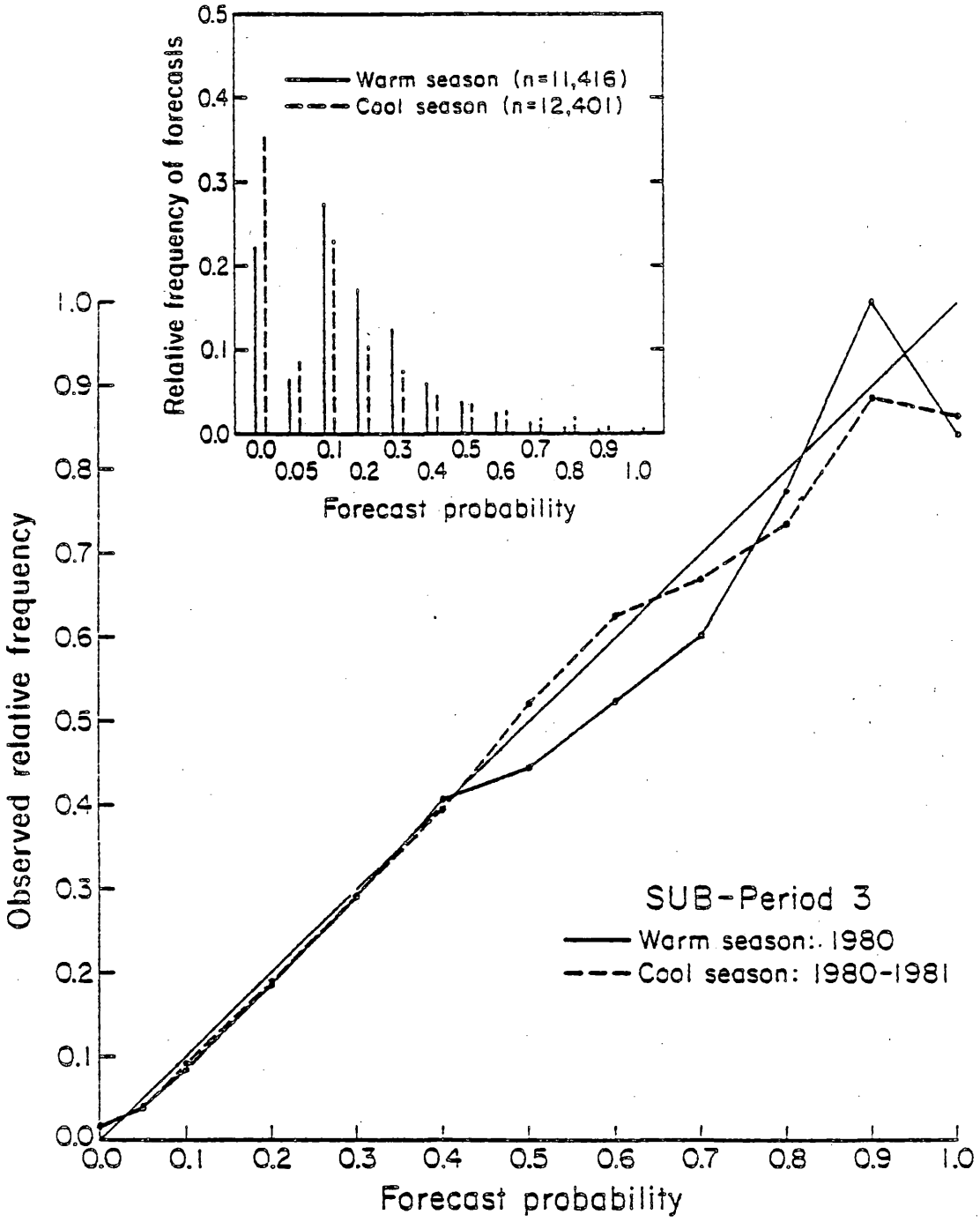

Figure 2c. Reliability diagram for third-period (36-48 hours) subjective PoP forecasts in 1980 warm season and 1980-1981 cool season for approximately 85 stations in conterminous U.S. Inset indicates relative frequency of use of probability values in respective seasons.

Table 5. Average Brier scores and skill scores based on Brier score of SUB probability of precipitation forecasts for three 12-hour periods (initial time 0000 GMT) in 1980 warm season and 1980-1981 cool season for approximately 85 stations in the conterminous U.S. (Maglaras et al., 1981; Schwartz et al., 1981).

	Warm season: 1980			Cool season: 1980-1981		
Period	Number of forecasts	Average Brier score[a]	Skill score (%)	Number of forecasts	Average Brier score[a]	Skill score (%)
1	11,489	0.0966	28.9	12,402	0.0733	49.5
2	11,493	0.1041	20.2	12,392	0.0926	33.4
3	11,416	0.1135	15.3	12,401	0.0946	34.2

[a]Average Brier score reported here is one-half of original Brier score.

forecasts are presented in Table 5. These results are similar to the results for the objective PoP forecasts (see Table 2); that is, skill generally decreases as a function of lead time and skill scores are higher in the cool season than in the warm season. Comparison of the skill scores for the subjective and objective PoP forecasts reveals that the subjective forecasts are more skillful for the first period, but that differences in skill for the second and third periods are quite small. It appears that (overall) the forecasters are able to improve appreciably upon the MOS guidance forecasts only in the first period.

Operational programs similar to the PoP program in the U.S. now exist in several countries. For example, meteorologists at the Royal Netherlands Meteorological Institute (KNMI) have made dry-period probability forecasts (the event of interest is the complement of the event of concern in PoP forecasts) for De Bilt on an operational basis since 1971 (Daan and Murphy, 1982). Moreover, the quality of these forecasts is comparable to the quality of the NWS precipitation probability forecasts. The KNMI dry-period forecasts are now available to the general public in The Netherlands via a teletext system. In Canada, the Atmospheric Environment Service (AES) initiated a nationwide precipitation probability forecasting program in July 1982 (Grimes, 1982). It is of interest to note that this program was undertaken (in part) in response to the general public's requests for PoP forecasts and that the initial results of this AES activity are quite encouraging. In addition to these operational programs, extensive experimental/operational trials involving PoP forecasts have been conducted in a number of countries, including Australia (Mason, 1976) and Finland (Saarikivi, 1980).

In Section 4.1, the method of successive subdivisions was described as a means of obtaining variable-width credible interval forecasts for continuous variables such as maximum and/or minimum temperatures. This approach has been used by NWS forecasters in several experiments (e.g., Peterson et al., 1972; Murphy and Winkler, 1974; Winkler and Murphy, 1979), and some results from two of these experiments are presented in Table 6. The data in this table relate to forecasts of maximum and minimum temperatures made at Denver, Colorado and Milwaukee, Wisconsin during the periods 1972-73 and 1974-75, respectively (see references for additional details). The reliability of these subjective credible interval forecasts can be evaluated by comparing the observed relative frequencies of temperatures below, in, and above the intervals with the probabilities of the intervals (the latter are 25.0%, 50.0%, and 25.0% for the 50% intervals and 12.5%, 75.0%, and 12.5% for the 75% intervals). Comparison of the relative frequencies and probabilities indicates that these credible interval forecasts are very reliable. Climatological credible interval forecasts were used as a standard of reference, and comparison of the reliability of the climatological forecasts and the subjective forecasts reveals that the latter are more reliable than the former. Moreover, the precision of the credible interval forecasts is also of consider-

Table 6. Relative frequency of occurrence of observed temperatures below interval (BI), in interval (II), and above interval (AI) for (a) subjective variable-width credible interval forecasts formulated in Denver and Milwaukee experiments (Murphy and Winkler, 1974; Winkler and Murphy, 1979) and (b) climatological forecasts corresponding to subjective forecasts.

		Relative frequency of observed temperatures (%)						Average width (standard deviation of width) (°F)	
		50% intervals			75% intervals				
Experiment	Number of forecasts	BI	II	AI	BI	II	AI	50% intervals	75% intervals
(a) Subjective forecasts									
Denver	132	25.8	45.5	28.8	10.6	73.5	15.9	6.2(1.3)	11.7(2.2)
Milwaukee	432	18.1	53.9	28.0	8.1	79.4	12.5	5.9(2.0)	10.1(3.1)
(b) Climatological forecasts									
Denver	132	31.1	44.7	24.2	18.9	65.2	15.9	14.8(4.2)	24.2(5.7)
Milwaukee	432	17.8	56.9	25.2	8.1	81.7	10.2	14.5(3.9)	23.7(4.9)

able interest; precision is measured here by the (average) width of the forecasts. Examination of the results in Table 6 indicates that the average width of the subjective credible interval forecasts is less than one-half the average width of the climatological forecasts. Thus, the variable-width temperature forecasts formulated in the Denver and Milwaukee experiments are both reliable and precise. It should be noted that, unlike the case of the forecasters who make the PoP forecasts, these forecasters had no prior experience making credible interval forecasts and they received no objective probabilistic temperature forecasts as guidance.

Subjective probability forecasts of wind speed, visibility, and precipitation events have been made on an experimental basis by KNMI forecasters at Zierikzee since 1980 (Daan and Murphy, 1982; Murphy and Daan, 1984). The forecasts specify the probabilities that certain critical threshold values of the respective variables will be exceeded or not exceeded in six-hour periods (see references for details). Evaluation of the results of the first year of the experiment reveals a considerable amount of overforecasting (Daan and Murphy, 1982); that is, forecast probabilities tended to exceed observed relative frequencies for all events. The forecasters were provided with extensive feedback at the beginning of the second year of the experiment, and the results for the two periods have been compared in a recent paper (Murphy and Daan, 1984). The reliability of the wind speed forecasts for the first and second years of the Zierikzee experiment is depicted in Figure 3 (reliability diagrams for the visibility and precipitation forecasts are similar and are omitted to conserve space). Moderate overforecasting occurred for the wind speed events for high probability values in year 1, but the feedback (and experience) appears to have reduced the overforecasting appreciably in year 2. The inset indicates that the frequency of use of the various probability values is similar in the two years, although probabilities in the highest range (0.90-1.00) were used less frequently in the second year.

Further evaluation of the subjective probabilistic wind speed, visibility, and precipitation forecasts formulated in the first and second years of the Zierikzee experiment is based upon the ranked probability score (RPS), and the results are presented in Table 7. The RPS itself is partitioned into three terms - variance of observations (UNC), reliability (REL), and resolution (RES) (with RPS = UNC + REL - RES) - and a skill score is computed based on this scoring rule (see Chapter 10 for definitions of these evaluation measures). The results indicate substantial improvement in reliability and skill from year 1 to year 2, whereas differences in resolution between the first and second years of the experiment are quite small. The increases in reliability and skill can be attributed primarily to the feedback provided to the forecasters in October 1981 (see Murphy and Daan, 1984). With regard to the differences in skill between the wind speed and visibility forecasts and the precipitation forecasts, it should be noted that the former are of much greater concern than the latter to the forecasters at Zierikzee (see Daan and Murphy, 1982). Finally,

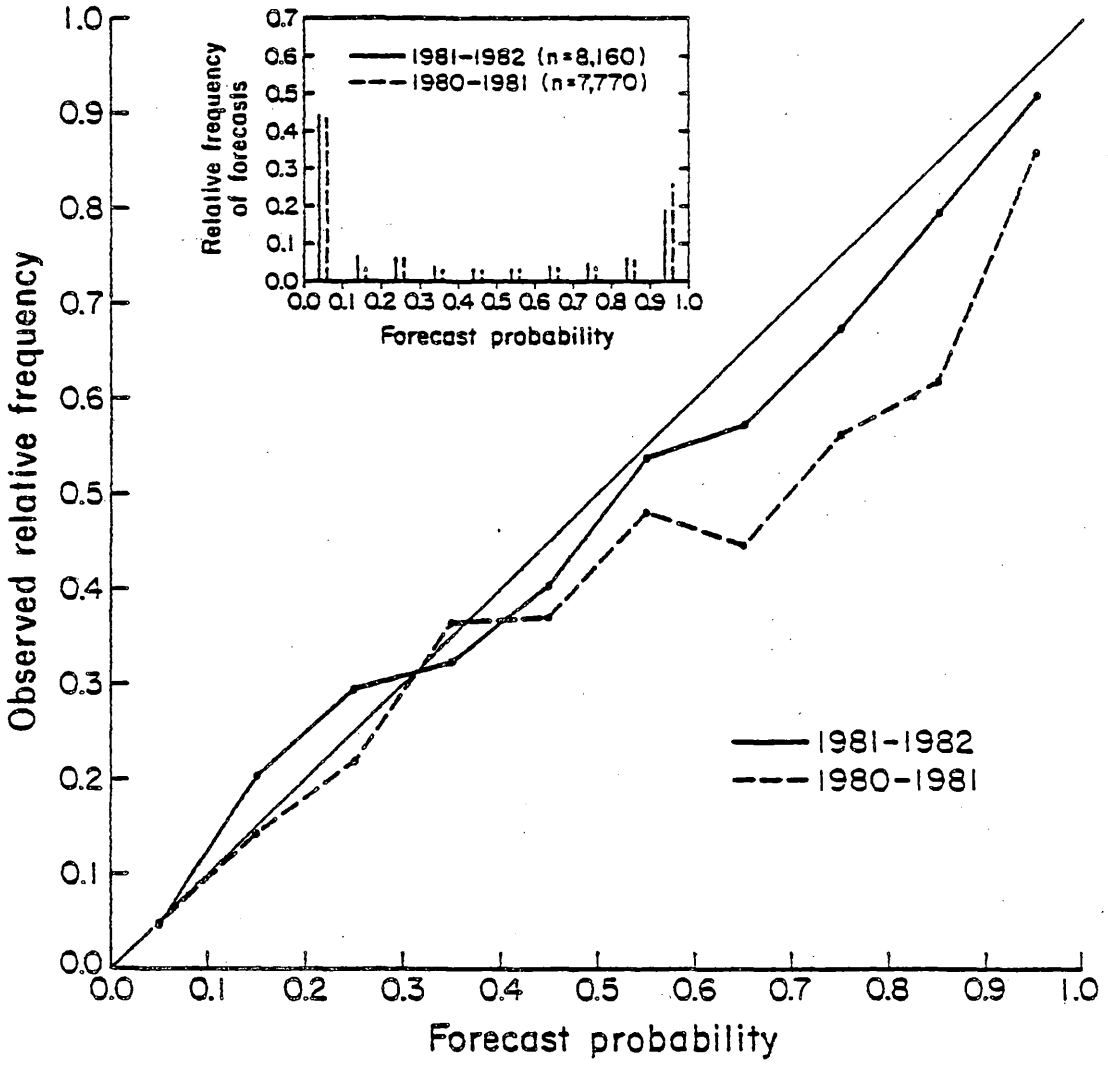

Figure 3. Reliability diagram for subjective probabilistic wind speed forecasts formulated in first and second years of Zierikzee experiment. Inset indicates relative frequency of use of probability values in respective periods.

Table 7. Partition of average ranked probability score and skill score based on ranked probability score for subjective probability forecasts of wind speed, visibility, and precipitation forecasts formulated in first and second years of Zierikzee experiment (Daan and Murphy, 1982; Murphy and Daan, 1984). Figures for first year are given in parentheses.

			Partition of ranked probability score			
Element	Number of thresholds	Number of probabilities	Variance of observations	Reliability	Resolution	Skill score (%)
Wind speed	3	8160(7770)	0.148(0.145)	0.005(0.016)	0.043(0.036)	25.4(13.9)
Visibility	3	8160(7770)	0.138(0.141)	0.007(0.018)	0.038(0.035)	22.4(12.4)
Precipitation	2	5440(5140)	0.122(0.143)	0.017(0.056)	0.018(0.021)	0.5(-24.7)

the forecasters at Zierikzee had no experience in probability forecasting prior to the start of the experiment, and they did not receive any objective probabilistic or categorical forecasts as guidance during the first two years of this program.

In addition to these experiments, several other experimental subjective probability forecasting programs have been conducted in the U.S. and elsewhere. For a review of the early history and developments in subjective probability forecasting in meteorology, see Murphy and Winkler (1984). More recently (i.e., since 1960), such programs in the U.S. have included probabilistic forecasting "games" conducted in conjunction with synoptic laboratory courses at several universities (e.g., Sanders, 1963, 1973, 1979; Bosart, 1975) and studies in which NWS forecasters have formulated probability forecasts on an experimental or quasi-operational basis. With regard to the latter, point and area precipitation probability forecasts have been prepared in several experiments (Winkler and Murphy, 1976; Murphy and Winkler, 1977; Murphy, 1978); probabilistic quantitative precipitation forecasts have been formulated for several locations in Texas (Murphy et al., 1982); forecasts of the probability that minimum temperatures will fall below 28°F have been prepared for fruit growers in New Mexico (Gregg, 1977); and probabilistic tornado forecasts have been made by forecasters at the NWS National Severe Storms Forecast Center (Murphy and Winkler, 1982). Recent probability forecasting experiments in other countries include the work of Mason (1976) in Australia and Saarikivi (1980) in Finland.

In this chapter we have been concerned primarily with short-range forecasts (i.e., forecasts of day-to-day weather events). However, monthly and seasonal forecasts of temperature and precipitation have been formulated subjectively (with objective inputs) by the NWS on an operational basis for more than two decades, and these forecasts are now provided to subscribers in a new probabilistic format (U.S. Department of Commerce, 1982). Specifically, the forecasts consist of probabilities for three categories of either average temperature or total precipitation in the respective periods (i.e., month or season). The categories are defined in such a manner that the historical climatological probabilites associated with these ranges of temperature or precipitation values are equal to 0.30, 0.40, and 0.30, respectively. In the context of forecasting, the probability of the central category ("near normal" in the case of temperature and "moderate" in the case of precipitation) is held constant at 0.40 on all occasions. Thus, each temperature or precipitation forecast can be expressed in the form of a single probability value. Evaluation of the NWS monthly and seasonal forecasts in their previous format reveals modest skill for temperature and only marginal skill for precipitation. The number of forecasts expressed in the new probabilistic format is still too small to support a systematic evaluation of their reliability and skill.

5. COMMUNICATION OF UNCERTAINTY IN WEATHER FORECASTS

In Sections 3 and 4 we described objective and subjective methods of probability forecasting and presented some results of operational and experimental probability forecasting programs. These results demonstrate that reliable and skillful probability forecasts can be formulated by both objective and subjective procedures. Nevertheless, with the exception of the precipitation probability forecasting program in the U.S. (and similar programs in a few other countries), probability forecasts generally are not provided to the general public and specific users on a regular basis. In view of the many successful operational and experimental probability forecasting programs and the fact that the value of probability forecasts generally exceeds the value of traditional categorical forecasts (see Section 2), the failure of the meteorological community to make greater use of probabilities in public (and other) weather forecasts is particularly surprising. Current resistance to the introduction of probabilities into weather forecasts appears to be based largely on issues related to the mode of expression of uncertainty in forecasts and the understanding of probability forecasts, and these issues are treated briefly in this section. The section concludes with a short discussion of the relative value of probabilistic and categorical forecasts.

Uncertainty can be described qualitatively in terms of words or quantitatively in terms of numbers. Traditionally, uncertainty in weather forecasts has been expressed in terms of verbal modifiers such as "chance" or "likely." Are these verbal expressions of uncertainty interpreted by the general public in an appropriate and consistent manner? This question has been addressed in several studies related to terminology in weather forecasts (e.g., Bickert, 1967; Abrams, 1971; Rogell, 1972; McBoyle, 1974) and in laboratory experiments conducted by behavioral psychologists (e.g., Lichtenstein and Newman, 1967; Budescu and Wallsten, 1981; Beyth-Marom, 1982). These investigators all reached the same general conclusion - a large amount of variability exists when individuals are asked to assign numerical values to such verbal expressions and the amount of overlap among terms is substantial. To eliminate this overlap, it would be necessary to limit the forecaster's vocabulary to a very small set of "distinct" words (e.g., always, often, sometimes, never), thereby severely restricting the amount of information concerning uncertainty that could be transmitted to users. Budescu and Wallsten (1981) conclude that "... communication from one person to another regarding degrees of uncertainty would be degraded by the use of probabilistic phrases rather than numbers." Unlike verbal expressions of uncertainty, probabilities can provide a precise and unambiguous description of the uncertainty in weather forecasts.

A related issue involves the interpretation by the general public of probabilistic weather forecasts such as the PoP forecasts. In this regard, selected individuals have been asked in several studies to interpret a PoP forecast such as "the probability of precipitation

today is 30%" (e.g., Bickert, 1967; Scoggins and Vaughn, 1971; Rogell, 1972). The results of these studies reveal that many individuals misinterpret PoP forecasts. However, it is important to distinguish between misinterpretation of the event (e.g., precipitation in an area instead of precipitation at a point) and misinterpretation of the probability associated with the event. In a recent study, Murphy et al. (1980) discovered that misinterpretation of PoP forecasts consists almost entirely of event misinterpretation rather than probability misinterpretation (the questions in the earlier studies were concerned exclusively with event misinterpretation). Moreover, this latter study revealed that a comparable amount of event misinterpretation exists for both categorical and probabilistic forecasts of precipitation occurrence. Thus, the misinterpretations of PoP forecasts cannot be attributed to the presence of probabilities in these forecasts. In fact, in view of the ambiguities associated with the interpretation of verbal expressions of uncertainty, the introduction of probabilities into public forecasts of precipitation occurrence actually may have led to a reduction in the overall level of misinterpretation of such forecasts.

The reactions of members of the general public to the precipitation probability forecasting program and their preferences regarding the mode of expression of uncertainty in forecasts are also of interest here. The PoP program initially encountered some resistance from both the forecasters and the public, but it is now generally agreed that these probabilities are an important and integral part of public weather forecasts in the U.S. (e.g., Bickert, 1967; American Telephone and Telegraph Company, 1971). Moreover, a recent nationwide statistical survey of 1300 members of the general public reveals that 70 percent of the participants prefer numerical probabilities to verbal modifiers as descriptors of uncertainty in forecasts of precipitation occurrence (M.S.I. Services, Inc., 1981). In addition, requests by the public in Canada for precipitation probability forcasts played a major role in the recent decision by the Atmospheric Environment Service to initiate a nationwide PoP program in July 1982 (Grimes, 1982).

In Section 2 we indicated that the value of probabilistic forecasts generally exceeds the value of categorical forecasts (e.g., see Thompson, 1962; Murphy, 1977; Krzysztofowicz, 1983). It should be noted that the studies on which this conclusion is based usually assume that the users of the forecasts process information and make decisions in a rational and optimal manner. Such an assumption may not be consistent with the information processing and decision-making procedures employed by most members of the general public or even by many specific users. However, Krzysztofowicz (1983) recently has compared the value of probabilistic and categorical forecasts for both optimal and suboptimal users and has concluded that the differences in value between the two types of forecasts are likely to be greater for suboptimal users than for optimal users. This result provides additional support for the use of probabilities in public weather forecasts.

ACKNOWLEDGMENTS

The preparation of this chapter was supported in part by the National Science Foundation (Atmospheric Sciences Division) under grant ATM-8004680. The author would like to thank the following staff members at the NWS Techniques Development Laboratory (Silver Spring, MD) for providing data and/or information on which the indicated tables and figures are based: G. M. Carter (Table 1); J. R. Bocchieri (Figures 1 and 2 and Table 3); and K. F. Hebenstreit (Table 4).

REFERENCES

Abrams, E. 1971: Problems in the communication of routine weather information to the public. University Park, Pennsylvania State University, Department of Meteorology, M.S. Thesis, 121 pp.

Air Weather Service, 1978: Probability forecasting: a guide for forecasters and staff weather officers. Scott Air Force Base, IL, U.S. Air Force, AWS Pamphlet 105-51, seven chapters and attachments (paged separately).

Allen, G., 1982: Probability judgment in weather forecasting. Preprints, Ninth Conference on Weather Forecasting and Analysis (Seattle). Boston, MA, American Meteorological Society, pp. 1-6.

American Telephone and Telegraph Company, 1971: Weather announcement study. New York, NY, American Telephone and Telegraph Company, Market and Service Plans Department, Report, 28 pp.

Balzer, K., 1980: Automated short-range area probability forecasts of thunderstorms, using the decision-tree method and observed temp part A data. Preprints, Symposium on Probabilistic and Statistical Methods in Weather Forecasting (Nice). Geneva, Switzerland, World Meteorological Organization, pp. 453-458.

Belville, J. D., and G. A. Johnson, 1982: The role of decision trees in weather forecasting. Preprints, Ninth Conference on Weather Forecasting and Analysis (Seattle). Boston, MA, American Meteorological Society, pp. 7-11.

Bermowitz, R. J., and E. A. Zurndorfer, 1979: Automated guidance for predicting quantitative precipitation. Monthly Weather Review, 107, 122-128.

Beyth-Marom, R., 1982: How probable is probable? A numerical translation of verbal probability expressions. Journal of Forecasting, 2, 257-269.

Bickert, C. von E., 1967: A study of the understanding and use of probability of precipitation forecasts in two major cities. Denver, CO, Unversity of Denver, Denver Research Institute, Report (Review Draft), 21 pp. plus four appendices.

Bocchieri, J. R., 1979a: A new operational system for forecasting precipitation type. Monthly Weather Review, 107, 637-649.

Bocchieri, J. R., 1979b: The use of LFM output for automated prediction of heavy snow. Preprints, Fourth Conference on Numerical Weather Prediction (Silver Spring). Boston, MA, American Meteorological Society, pp. 77-81.

Bosart, L. F., 1975: SUNYA experimental results in forecasting daily temperature and precipitation. Monthly Weather Review, 103, 1013-1020.

Brelsford, W. M., and R. H. Jones, 1967: Estimating probabilities. Monthly Weather Review, 95, 570-576.

Brier, G. W., 1946: A study of quantitative precipitation forecasting in the TVA basin. Washington, D.C., U.S. Weather Bureau, Research Paper No. 26, 40 pp.

Budescu, D. V., and T. S. Wallsten, 1981: Intra and intersubject consistency in comparing probability phrases. Chapel Hill, University of North Carolina, Psychometric Laboratory, Research Memorandum No. 49, 25 pp.

Carter, G. M., and H. R. Glahn, 1976: Objective prediction of cloud amount based on Model Output Statistics. Monthly Weather Review, 104, 1565-1572.

Charba, J. P., 1979: Two to six hour severe local storm probabilities: an operational forecasting system. Monthly Weather Review, 107, 268-282.

Charba, J. P., and W. H. Klein, 1980: Skill in precipitation forecasting in the National Weather Service. Bulletin of the American Meteorological Society, 61, 1546-1555.

Cooke, W. E., 1906a: Forecasts and verifications in Western Australia. Monthly Weather Review, 34, 23-24.

Cooke, W. E., 1906b: Weighting forecasts. Monthly Weather Review, 34, 274-275.

Daan, H., and A. H. Murphy, 1982: Subjective probability forecasting in The Netherlands: some operational and experimental results. Meteorologische Rundschau, 35, 99-112.

Dickey, W. W., 1949: Estimating the probability of a large fall in temperature at Washington, D.C. Monthly Weather Review, 77, 67-78.

Enger, I., J. A. Russo, Jr., and E. L. Sorenson, 1964: A statistical approach to 2-7 hr prediction of ceiling and visibility, volumes I and II. Hartford, CT, Travelers Research Center, Inc., Contract No. Cwb-10704, 48 pp. and 195 pp., respectively.

Epstein, E. S., 1969: Stochastic dynamic prediction. Tellus, 21, 739-759.

Fischhoff, B., and D. MacGregor, 1982: Subjective confidence in forecasts. Journal of Forecasting, 1, 155-172.

Fix, C., and J. L. Hodges, Jr., 1951: Discriminatory analysis, non-parametric discrimination: consistency properties. Randolph Field, TX, USAF, School of Aviation Medicine, Report No. 4, 19 pp.

Freeman, M. H., 1961: A graphical method of objective forecasting derived by statistical techniques. Quarterly Journal of the Royal Meteorological Society, 87, 393-400.

Glahn, H. R., 1964: An application of adaptive logic to meteorological prediction. Journal of Applied Meteorology, 3, 718-725.

Glahn, H. R., 1980: Methods and accuracy of statistical weather forecasting in the United States. Preprints, Symposium on Probabilistic and Statistical Methods in Weather Forecasting (Nice). Geneva, Switzerland, World Meteorological Organization, pp. 387-395.

Globokar, F. T., 1974: Computerized ceiling and visibility forecasts. Preprints, Fifth Conference on Weather Forecasting and Analysis (St. Louis). Boston, MA, American Meteorological Society, pp. 228-233.

Gregg, G. T., 1977: Probability forecasts of a temperature event. National Weather Digest, 2, No. 2, 33-34.

Grimes, D., 1982: A preliminary verification of the Pacific Weather Centre's probability of precipitation forecasts. Vancouver, B.C., Canada, Atmospheric Environment Service, Pacific Region Technical Notes 82-013, 8 pp.

Gringorten, I. I., 1966: A stochastic model of the frequency and duration of weather events. Journal of Applied Meteorology, 5, 606-624.

Gringorten, I. I., 1971: Modelling conditional probability. Journal of Applied Meteorology, 10, 646-657.

Hampton, J. M., P. G. Moore, and H. Thomas, 1973: Subjective probability and its measurement. Journal of the Royal Statistical Society, Series A, 136, 21-42.

Hughes, L. A., 1980: Probability forecasting - reasons, procedures, problems. Silver Spring, MD, U.S. Department of Commerce, NOAA, National Weather Service, Technical Memorandum NWS FCST 24, 84 pp.

Javelle, J. P., J. M. Veysseire, A. M. Calvayrac, F. Duvernet, and G. Der Megreditchian, 1980: Operational forecasting of local weather parameters through statistical interpretation of NWP using "perfect prog" methods. Preprints, Symposium on Probabilistic and Statistical Methods in Weather Forecasting (Nice). Geneva, Switzerland, World Meteorological Organization, pp. 423-429.

Jensenius, J. S., and G. M. Carter, 1979: Automated forecasts of agricultural weather elements. Preprints, Fourteenth Conference on Agriculture and Forest Meteorology (Minneapolis). Boston, MA, American Meteorological Society, pp. 42-44.

Jones, R. H., 1968: A nonlinear model for estimating probabilities of k events. Monthly Weather Review, 96, 383-384.

Kruizinga, S., and A. H. Murphy, 1983: Use of an analogue procedure to formulate objective probabilistic temperature forecasts in The Netherlands. Monthly Weather Review, 111, 2244-2254.

Krzysztofowicz, R., 1983: Why should a forecaster and a decision maker use Bayes' theorem? Water Resources Research, 19, 327-336.

Lichtenstein, S., and J. R. Newman, 1967: Empirical scaling of common verbal phrases associated with numerical probabilities. Psychonomic Sciences, 9, 563-564.

Lichtenstein, S., B. Fischhoff, and L. D. Phillips, 1982: Calibration of probabilities: the state of the art to 1980. Judgment under Uncertainty: Heuristics and Biases (D. Kahneman, P. Slovic, and A. Tversky, Editors). London, Cambridge University Press, pp. 306-334.

Lindley, D. V., 1965: Introduction to Probability and Statistics from a Bayesian Viewpoint: Part I. Probability. London, Cambridge University Press, 259 pp.

Lowry, D. A., and H. R. Glahn, 1976: An operational model for forecasting probability of precipitation - PEATMOS PoP. Monthly Weather Review, 104, 221-232.

Maglaras, G. J., J. P. Dallavalle, K. F. Hebenstreit, G. H. Hollenbaugh, B. E. Schwartz, and D. J. Vercelli, 1981: Comparative verification of guidance and local aviation/public weather forecasts - No. 10 (April 1980 - September 1980). Silver Spring, MD, U.S. Department of Commerce, NOAA, National Weather Service, TDL Office Note 81-7, 61 pp.

Martin, D. E., 1972: Climatic presentations for short-range forecasting based on event occurrence and reoccurrence profiles. Journal of Applied Meteorology, 11, 1212-1223.

Mason, I. B., 1976: Weather forecasts as subjective probability statements: a report on an experiment. Canberra, Australia, Bureau of Meteorology, Regional Forecast Centre, unpublished manuscript, 36 pp.

Mason, I. B., 1977: Objective probabilities for fog at Canberra Airport using Bayesian aggregation and principal components. Australian Meteorological Magazine, 25, 121-134.

McBoyle, G. R., 1974: Public response to weather terminology in the Kitchener-Waterloo area. Climatological Bulletin (McGill University), No. 15, pp. 11-29.

Miller, R. G., 1962: Statistical prediction by discriminant analysis. Meteorological Monographs, 4, No. 25, 54 pp.

Miller, R. G., 1964: Regression estimation of event probabilities. Hartford, CT, Travelers Research Center, Contract Cwb-10704, Technical Report No. 1, 153 pp.

M.S.I. Services, Inc., 1981: Public requirements for weather information and attitudes concerning weather service. Washington, D.C., M.S.I. Services, Inc., Technical Report, 46 pp. plus three appendices.

Murphy, A. H., 1977: The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situation. Monthly Weather Review, 105, 803-816.

Murphy, A. H., 1978: On the evaluation of point precipitation probability forecasts in terms of areal coverage. Monthly Weather Review, 106, 1680-1686.

Murphy, A. H., and H. Daan, 1984: Impacts of feedback and experience on the quality of subjective probability forecasts: comparison of results from the first and second years of the Zierikzee experiment. Monthly Weather Review, 112, 413-423.

Murphy, A. H., W.-R. Hsu, and R. L. Winkler, 1982: Subjective probabilistic quantitative precipitation forecasts: some experimental results. Preprints, Ninth Conference on Weather Forecasting and Analysis (Seattle). Boston, MA, American Meteorological Society, pp. 94-100.

Murphy, A. H., S. Lichtenstein, B. Fischhoff, and R. L. Winkler, 1980: Misinterpretations of precipitation probability forecasts. Bulletin of the American Meteorological Society, 61, 695-701.

Murphy, A. H., and R. L. Winkler, 1971: Forecasters and probability forecasts: some current problems. Bulletin of the American Meteorological Society, 52, 239-247.

Murphy, A. H., and R. L. Winkler, 1974: Credible interval temperature forecasting: some experimental results. Monthly Weather Review, 102, 784-794.

Murphy, A. H., and R. L. Winkler, 1977: Experimental point and area precipitation probability forecasts for a forecast area with significant local effects. Atmosphere, 15, 61-78.

Murphy, A. H., and R. L. Winkler, 1982: Subjective probabilistic tornado forecasting: some experimental results. Monthly Weather Review, 110, 1288-1297.

Murphy, A. H., and R. L. Winkler, 1984: Probability forecasting in meteorology. Journal of the American Statistical Association, 79, in press.

Peterson, C. R., K. J. Snapper, and A. H. Murphy, 1972: Credible interval temperature forecasts. Bulletin of the American Meteorological Society, 53, 966-970.

Raiffa, H., 1968: Decision Analysis: Introductory Lectures on Choices under Uncertainty. Reading, MA, Addison-Wesley, 309 pp.

Reap, R. M., and D. S. Foster, 1979: Automated 12-36 hour probability forecasts of thunderstorms and severe local storms. Journal of Applied Meteorology, 18, 1304-1315.

Roberts, C. F., J. M. Porter, and G. F. Cobb, 1969: Report on Weather Bureau forecast performance 1967-68 and comparison with previous years. Silver Spring, MD, U.S. Department of Commerce, ESSA, Weather Bureau, Technical Memorandum WBTM FCST 11, 44 pp.

Rogell, R. H., 1972: Weather terminology and the general public. Weatherwise, 25, 126-132.

Saarikivi, P., 1980: The effect of the temporal variability of precipitation on the quality of probability forecasts. Preprints, Symposium on Probabilistic and Statistical Methods in Weather Forecasting (Nice). Geneva, Switzerland, World Meteorological Organization, pp. 229-234.

Sanders, F., 1963: On subjective probability forecasting. Journal of Applied Meteorology, 2, 191-201.

Sanders, F., 1973: Skill in forecasting daily temperature and precipitation: some experimental results. Bulletin of the American Meteorological Society, 54, 1171-1179.

Sanders, F., 1979: Trends in skill of daily forecasts of temperature and precipitation, 1966-78. Bulletin of the American Meteorological Society, 60, 763-769.

Schmidt, R. C., 1951: A method of forecasting occurrence of winter precipitation two days in advance. Monthly Weather Review, 79, 81-95.

Schwartz, B. E., J. R. Bocchieri, G. M. Carter, J. P. Dallavalle, G. H. Hollenbaugh, G. J. Maglaras, and D. J. Vercelli, 1981: Comparative verification of guidance and local aviation/public weather forecasts - No. 11 (October 1980 - March 1981). Silver Spring, MD, U.S. Department of Commerce, NOAA, National Weather Service, TDL Office Note 81-10, 77 pp.

Scoggins, J. R., and W. W. Vaughn, 1971: How some nonmeteorological professionals view meteorology and weather forecasting. Bulletin of the American Meteorological Society, 52, 974-979.

Spetzler, C. S., and C.-A. S. Staël von Holstein, 1975: Probability encoding in decision analysis. Management Science, 22, 340-358.

Staël von Holstein, C.-A. S., 1970: Assessment and Evaluation of Subjective Probability Distributions. Stockholm, Sweden, Stockholm School of Economics, Economic Research Institute, 225 pp.

Thompson, J. C., 1950: A numerical method for forecasting rainfall in the Los Angeles area. Monthly Weather Review, 78, 113-124.

Thompson, J. C., 1962: Economic gains from scientific advances and operational improvements in meteorological prediction. Journal of Applied Meteorology, 1, 13-17.

Tversky, A., and D. Kahneman, 1974: Judgment under uncertainty: heuristics and biases. Science, 185, 1124-1131.

U.S. Department of Commerce, 1982: Monthly and seasonal weather outlook. Washington, D.C., NOAA, National Weather Service, Climate Analysis Center, Volume 36, Number 14, 8 pp. (Published twice a month and available from U.S. Government Printing Office.)

Wahl, E. W., and collaborators, 1952: The construction and application of contingency tables in weather forecasting. Cambridge, MA, U.S. Air Force Cambridge Research Laboratories, Surveys in Geophysics No. 19, 37 pp.

Wallsten, T. S., and D. V. Budescu, 1983: Encoding subjective probabilities: a psychological and psychometric review. Management Science, 29, 151-173.

Wilson, L. J., and N. Yacowar, 1980: Statistical weather element forecasting in the Canadian Weather Service. Preprints, Symposium on Probabilistic and Statistical Methods in Weather Forecasting (Nice). Geneva, Switzerland, World Meteorological Organization, pp. 401-406.

Winkler, R. L., 1967: The assessment of prior distributions in Bayesian analysis. Journal of the American Statistical Association, 62, 776-800.

Winkler, R. L., 1972: Introduction to Bayesian Inference and Decision. New York, Holt, Rinehart and Winston, 563 pp.

Winkler, R. L., and A. H. Murphy, 1968: "Good" probability assessors. Journal of Applied Meteorology, 7, 751-758.

Winkler, R. L., and A. H. Murphy, 1973: Information aggregation in probabilistic prediction. IEEE Transactions on Systems, Man, and Cybernetics, SME-3, 154-160.

Winkler, R. L., and A. H. Murphy, 1976: Point and area precipitation probability forecasts: some experimental results. Monthly Weather Review, 104, 86-95.

Winkler, R. L., and A. H. Murphy, 1979: The use of probabilities in forecasts of maximum and minimum temperatures. Meteorological Magazine, 108, 317-329.

Winkler, R. L., A. H. Murphy, and R. W. Katz, 1977: The consensus of subjective probability forecasts: are two, three, ... heads better than one? Preprints, Fifth Conference on Probability and Statistics in Atmospheric Sciences (Las Vegas). Boston, MA, American Meteorological Society, pp. 57-62.

Zurndorfer, E. A., J. R. Bocchieri, G. M. Carter, J. P. Dallavalle, D. B. Gilhousen, K. F. Hebenstreit, and D. J. Vercelli, 1979: Trends in verification scores for guidance and local aviation/public weather forecasts. Monthly Weather Review, 107, 799-811.